Remarks

Upon entry of the foregoing amendment, claims 24-39, 43-100, 105-139 and 141-156 will be pending in the application, with claims 24, 51, 75, 84, 105, 107, 109, 111, 121, 128, 137 and 149 being the independent claims. Claims 24, 51, 105, 107, 109, 111, 128, 137 and 154 have been amended taking the Examiner's comments into consideration. New claim 156 was added. This amendment introduces no new matter and entry thereof is respectfully requested.

Support for the amended claims can be found throughout the specification. Support for amended claims 24, 51, 105, 107, 109 and 111 can be found, *inter alia*, at page 9, lines 27-28; at page 25, lines 30-31; at page 64, lines 27-29; at page 65, lines 2-8; at page 87, lines 12-16; and at page 91, lines 8-9. Support for claim 128 can be found, *inter alia*, at page 23, lines 9-13. Support for claim 137 can be found, *inter alia*, at page 25, lines 28-29; and at page 26, lines 14-16. Support for claim 154 can be found, *inter alia*, at page 27, line 27, to page 28, line 4. Support for new claim 156 can be found, *inter alia*, at page 18, line 30, to page 19, line 9.

Based on the above amendments and the following remarks, Applicants respectfully request that the Examiner reconsider all outstanding objections and rejections and that they be withdrawn.

Objection to the Specification

The Examiner objected to the disclosure stating that "[t]he 'Brief Description of the Drawings' (pages 2 and 3), specifically for Figures 1A-1C and 4-6, needs to be amended to reflect the new labeling present in the formal drawings filed 2/22/01, including where panels

in the original figures are now given their own figure numbers." (Paper No. 29, at page 2.)

Applicants have amended the specification to comply with the Examiner's request. Thus, this objection has been rendered moot.

Rejections under 35 U.S.C. § 112

Claims 24-26, 28-29, 31-32, 34-35, 37-38, 43-53, 55-56, 58-59, 61-62, 64-65, 67-74, 79, 81-82, 105, 107, 109, 111, 113-120, 125, 127, 132, 134-135, 144, 146-147 and 149-155 were rejected under 35 U.S.C. § 112, first paragraph, for allegedly containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors, at the time the application was filed, had possession of the claimed invention. (Paper No. 29, at page 3.)

Regarding claims 24, 51, 105, 107 and 111 (and their dependent claims), the Examiner alleged that "[t]he specification does not teach using 'variant' polypeptides, such as those encoded by the broadly claimed polynucleotides, in order to make antibodies against the protein of SEQ ID NO:2." *Id.* Applicants respectfully disagree.

In contrast to the Examiner's assertions, the specification does teach using variant polypeptides, such as those encoded by the claimed polynucleotides, to make antibodies against the protein of SEQ ID NO:2. For example, the specification teaches that even if deletion of

> one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the protein will likely be retained when less than the majority of the residues of the protein are removed from the N-terminus or C-terminus.

The specification at page 16, lines 4-18. Hence, the specification teaches that a PDEF deletion variant can induce and/or bind antibodies. The specification goes on to teach that other variants include insertions, inversions, repeats and substitutions. *Id.*

In addition, the specification describes phenotypically silent amino acid substitutions (*see* page 16, lines 15-31). Furthermore, the genetic code is known. Thus, even though the claims do not require the nucleic acid to encode a polypeptide identical to SEQ ID NO:2, a polynucleotide with a nucleotide sequence different from that of SEQ ID NO:1 can still produce a polypeptide identical to that of SEQ ID NO:2 which can be used to induce antibodies that bind the protein of SEQ ID NO:2.

However, solely in an effort to advance prosecution, Applicants have amended the claims to recite that the nucleic acid encodes a polypeptide which binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2. Accordingly, withdrawal of this rejection is respectfully requested.

Regarding claims 46, 48, 70, 72, 79, 81, 116, 118, 125, 127, 132, 134, 144, 146, 149 and claims dependent therefrom, the Examiner alleged that "[t]he terms 'heterologous regulatory sequence' and 'regulatory elements' do not appear in the specification as originally filed, nor does any generic term equivalent to these terms" and "[i]t is unclear just what these terms are intended to convey. There is no clear support for these limitations in the original specification." (Paper No. 29, at page 4.) Applicants respectfully traverse this rejection.

First of all, the term "regulatory regions," which is a generic term equivalent to "regulatory sequence," and the term "regulatory elements" appear throughout the specification (*see* specification at page 1, lines 13-18; page 36, lines 18-21; and page 99, lines 3-8). In addition, there are many examples of regulatory/control sequences described

in the specification, including ribosome binding sites and initiator ATG, which are translational control sequences. *See also* Current Protocols in Molecular Biology, Volume 2, pp. 16.1.1-16.1.3; and 16.12.1-16.12.6 (Attachments A and B). Other examples include enhancers, promoters and polyadenylation signal sequences (*see* page 27, line 27, to page 28, line 4, of the specification), which are transcription regulatory elements (*see also* Current Protocols in Molecular Biology, Volume 2, pp. 16.13.1-16.13.7)(Attachment C). Additionally, the expression vectors disclosed as being part of the invention contain other regulatory sequences (*see, e.g.*, Pharmacia BioTech, BioDirectory, Section 5: cDNA, Cloning & Vectors, pp. 91-93, 1997)(Attachment D).

This large number of disclosed regulatory elements constitutes an adequate written description for transcriptional, translational and heterologous regulatory sequences and would be recognized as such by one of ordinary skill in the art. The specification contemplates use of any known sequences; thus, the invention is not limited to the particular elements disclosed. Thus, there is clear support for these terms in the specification. One skilled in the art reading the specification would immediately understand what the terms "heterologous regulatory sequence" and "regulatory elements" are intended to convey.

The specification also teaches, at page 54, lines 15-23, that the present invention includes an expression vector comprising phage operator and promoter elements operatively linked to a PDEF polynucleotide. The vector further contains an *E. coli* origin of replication, a T phage promoter sequence, two lac operator sequences, a Shine-Delgarno sequence and the lactose operon repressor gene. All of these sequences are examples of other types of regulatory sequences or elements. *See also* page 52, lines 21 to 30; page 56, lines 20 to 31; page 57, lines 1 to 15; page 59, lines 8-20; page 60, lines 7-14; page 93, lines 26-28; page

94, lines 17–18; and page 99, lines 17-19 and 24-26. Therefore, the specification conveys with reasonable clarity that Applicants possessed the claimed invention at the time the application was filed.

Furthermore, the absence of definitions or details for well-established terms or procedures should not be the basis of a rejection under 35 U.S.C. § 112, first paragraph, for lack of adequate written description. *See* 66 Fed. Reg. 1099, 1105 (Jan. 5, 2001)(inverse correlation between level of skill and knowledge and specificity of disclosure needed to satisfy written description; written description need only describe in detail what is new or not conventional.); and *Spectra-Physics, Inc. v. Coherent, Inc.*, 827 F.2d 1524, 1534 (Fed. Cir. 1987)("[a] patent need not teach, and preferably omits, what is well known in the art.") Accordingly, withdrawal of this rejection is respectfully requested.

The Examiner further alleged that claim 154 "does not limit the identity of the polypeptide to that encoded by the recombinant polypeptide, nor even that the cell used be able to transcribe or translate the recombinant polynucleotide" and "[c]laim 154 should be amended to provide some nexus between the 'polypeptide' recited in claim 154 and the '60 contiguous amino acids' recited in claim 149." (Paper No. 29, at page 5.)

Applicants respectfully disagree. However, in an effort to advance prosecution, Applicants have amended claim 154 as the Examiner suggested. Thus, withdrawal of this rejection is respectfully requested. Applicants point out that the amendment does not narrow the claim in any way since Applicants intended the claim to encompass the polynucleotide recited in claim 149.

The Examiner maintained the rejection of claims 137-139 and 141-148 under 35 U.S.C. § 112, first paragraph, as allegedly containing subject matter which was not described

in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention. (Paper No. 29, at page 5.) According to the Examiner,

> there is no apparent support for the generic embodiment instantly claimed with respect to "nucleotide sequence heterologous to SEQ ID NO:1" in the context of a fusion to nucleic acid encoding an epitope of PDEF. There is no evidence of record that such a generic embodiment was contemplated by the inventors at the time the invention was made.
>
> Applicant's arguments filed 2/7/01 have been fully considered but they are not persuasive. The amendment does not obviate the prior grounds of rejection since the claim still embraces a genus for which there is no description in the original specification, embracing embodiments where the "nucleotide sequence heterologous to SEQ ID NO:1" does not encode a "second protein" (see specification page 25, lines 28-29). The phrase "fused in frame" has no meaning in the context of the claim as written because there is no indication of any element contained in the "nucleotide sequence heterologous to SEQ ID NO:1" that has a "frame."

Id. at page 6. Applicants respectfully disagree.

The specification, *inter alia*, at page 26, lines 14-16, teaches that PDEF polypeptides, including fragments, and *specifically epitopes*, can be combined with parts of the constant domain of immunoglobulins (IgG), *e.g.*, a nucleotide sequence heterologous to SEQ ID NO:1, to form chimeric polypeptides (emphasis added).

Furthermore, the specification at page 46, lines 10-13, teaches that the PDEF promoter may be *fused* to a coding sequence to obtain prostate specific expression of the coding sequence (emphasis added). One skilled in the art reading the above disclosure would thus understand how to fuse an epitope in frame to a nucleotide sequence heterologous to SEQ ID NO:1, as recited in claim 137. Therefore, the meaning of the phrase "fused in frame" is clear in the context of the claim as written. However, in an effort to

advance prosecution, Applicants have amended claim 137 to indicate that the heterologous polynucleotide encodes a heterologous polypeptide. Thus, the nucleotide sequence heterologous to SEQ ID NO:1 encodes "a second protein." Accordingly, withdrawal of this rejection is respectfully requested. Applicants note that this amendment does not narrow the claims in any way since Applicants intended the claim to encompass a fusion protein.

The Examiner maintained the rejection of claims 24-26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43-53, 55, 56, 58, 59, 61, 62, 64, 65, 67-74, 105, 107, 109, 111 and 113-120 under 35 U.S.C. § 112, first paragraph, because allegedly "the specification, while being enabling for a 'nucleic acid' that encodes SEQ ID NO:2 or a fragment of SEQ ID NO:2 (as recited in the claims), does not reasonably provide enablement for polynucleotides that do not encode SEQ ID NO:2 or a recited fragment of SEQ ID NO:2." (Paper No. 29, at page 6.) According to the Examiner, "[t]he specification does not teach using 'variant' polypeptides, such as those encoded by the broadly claimed polynucleotides, in order to make antibodies against the protein of SEQ ID NO:2." *Id.* at page 7.

Applicants respectfully submit that the claims are not to methods of making antibodies. Instead, the claims are to nucleic acid variants that are 90% identical to nucleic acids that encode SEQ ID NO:2 or portions thereof. Methods of making such variants are well known and taught in the specification. *See* the specification, *inter alia*, at page 11, line 8, to page 17, line 21. The Examiner contends that the claims require raising antibodies against specified variants. Although the specification enables such a method, the claims do not require it. Whether such variant polypeptides bind to an antibody specific to SEQ ID NO:2 can be readily determined using an antibody raised against SEQ ID NO:2, not against the variant as the Examiner asserts is required by the claims.

Furthermore, it is well established that such variants would be useful to examine the protein of SEQ ID NO:2. For example, it is well established in the art of molecular biology that such variants can be routinely made and used in epitope-mapping studies. *See, e.g.*, Ikeda *et al.*, "Epitope mapping of anti-recA protein IgGs by region specified polymerase chain reaction mutagenesis," *J. Biol. Chem. 267*:6291-6296 (1992)(Attachment E). This is merely routine experimentation. Thus, the claims are fully enabled by the specification.

Moreover, as Applicants discussed above, the specification does teach that a polynucleotide with a sequence different from that of SEQ ID NO:1 can still produce a polypeptide identical to that of SEQ ID NO:2 which can be used to induce antibodies that bind the protein of SEQ ID NO:2. In addition, the specification at page 16, lines 4-18, describes deletion variants which can be used to induce antibodies that bind the polypeptide of SEQ ID NO:2. The ability of a deletion variant to induce and/or to bind antibodies which recognize the full-length protein will likely be retained when less than the majority of the residues of the protein are removed from the N-terminus or C-terminus. The specification goes on to teach that other variants include insertions, inversions, repeats and substitutions. Thus, in contrast to the Examiner's statement that "[f]or making antibodies that bind to PDEF, the specification (pages 24-25) teaches to use PDEF (SEQ ID NO:2) or an antigenic fragment of PDEF; no mention is made of using variant polypeptides for this purpose," (Paper No. 29, at page 9), the specification clearly teaches that PDEF variants can be used for making antibodies.

In addition, the experimentation required to make and use such polynucleotides requires little if any ingenuity and, similar to the situation in *In re Wands*, 8 U.S.P.Q.2d 1400 (C.A.F.C. 1988), is merely routine experimentation for one skilled in the art of molecular

biology. Thus, in contrast to the Examiner's assertions, the situation here meets the tests in *Fields v. Conover*, 170 U.S.P.Q. 276, 279 (C.C.P.A. 1971).

Furthermore, the Examiner has provided no evidence that one skilled in the art of molecular biology would be unable to make and use nucleic acid variants that are 90% identical to nucleic acids that encode SEQ ID NO:2 or portions thereof. Instead, the Examiner has merely recited case law to support his assertions.

However, solely in an effort to advance prosecution, Applicants have amended the claims to recite that the nucleic acid encodes a polypeptide which binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2. Thus, withdrawal of this rejection is respectfully requested.

The Examiner further rejected claims 24-39, 43-74 and 105-120 under 35 U.S.C. § 112, second paragraph, as allegedly being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as the invention. (Paper No. 29, at page 11.) According to the Examiner,

> [c]laims 24, 51, 105, 107 and 111 (and their dependent claims) recite the limitation that the nucleic acid encodes "a polypeptide which *generates* an antibody that *binds* a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2" (emphasis added). First, it is unclear what the term "generates" means in this context. . . . Second, the meets and bounds of the claim are unclear with respect to what "binds" is intended to mean in this context.

Id. at pages 11-12.

Applicants respectfully disagree. However, as indicated above in response to the 35 U.S.C. § 112, first paragraph, rejection, Applicants have amended the claims in an effort to advance prosecution. The amended claims do not recite "generates" and are clear with

respect to what "binds" means. Accordingly, withdrawal of this rejection is respectfully requested.

Rejections under 35 U.S.C. § 102

Claims 128-136 remained rejected under 35 U.S.C. § 102(a) as allegedly being anticipated by GenBank Accession No. AA662204. (Paper No. 29, at page 12.) According to the Examiner, "[t]he prior art polynucleotide includes codons for a DS dipeptide corresponding to SEQ ID NO:2 positions 283-284, respectively, mentioned above, which anticipates the claims when m and n are either 237 and 238, respectively, or 241 and 242, respectively." *Id.*

Applicants have amended claim 128 to recite that m is an integer from 2 to 236 and n is an integer from 243 to 335. Thus, the cited dipeptide is not encompassed by the claim. Support for this amendment can be found at pages 19-23 of the specification. Accordingly, this rejection has been rendered moot.

The Examiner further maintained the rejection of claims 128-136 under 35 U.S.C. § 102(b) as allegedly being anticipated by Chen *et al.*, *Dev. Biol. 151*:176-191 (1992). (Paper No. 29, at page 13.)

According to the Examiner, "[t]he ets-4 polypeptide has a DS dipeptide at residues 60-61, which anticipates the claims when m and n are either 237 and 238, respectively, or 241 and 242, respectively." *Id.*

As indicated above, Applicants have amended claim 128 to recite that m is an integer from 2 to 236 and n is an integer from 243 to 335. Thus, the cited dipeptide is not encompassed by the claim. Accordingly, this rejection has been rendered moot. Applicants emphasize that the value "m is an integer from 2 to 236" and "n is an integer from 243 to

335" is not critical to the invention. Thus, this amendment should not be construed as a surrender of equivalent embodiments wherein m is more than 236 or n is less than 243 that do not read on the cited references.

Double patenting

The Examiner again advised Applicants that should claims 27, 30, 33, 36 and 39 be found allowable, claims 54, 57, 60, 63 and 66 would be objected to under 37 C.F.R. § 1.75 as being a substantial duplicate thereof. (Paper No. 29, at page 13.)

In response, Applicants note that they are entitled to claim the invention using multiple claims, so long as the claim set as a whole clearly defines the subject matter of the invention. *See* M.P.E.P. § 2173.05(n). Further, Applicants note that claims 27, 30, 33, 36 and 39 and 54, 57, 60, 63 and 66 are not duplicative, because claims 27, 30, 33, 36 and 39 are drawn to an isolated polynucleotide comprising a ***nucleic acid at least 90% identical to a reference nucleic acid*** encoding certain amino acids of SEQ ID NO:2, whereas claims 54, 57, 60, 63 and 66 are drawn to an isolated polynucleotide comprising a nucleic acid encoding an ***amino acid sequence at least 90% identical to a reference amino acid sequence*** encoding certain amino acids of SEQ ID NO:2. Thus, the claims are of different scope. Accordingly, withdrawal of this objection is respectfully requested.

Allowable Subject Matter

The indication that claims 75-78, 80, 83-100, 121-124 and 126 are allowed is noted and appreciated by Applicants.

Conclusion

All of the stated grounds of objection and rejection have been properly traversed, accommodated, or rendered moot. Applicants therefore respectfully request that the Examiner reconsider all presently outstanding objections and rejections and that they be withdrawn. Applicants believe that a full and complete reply has been made to the outstanding Office Action and, as such, the present application is in condition for allowance. If the Examiner believes, for any reason, that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned at the number provided.

Prompt and favorable consideration of this Amendment and Reply is respectfully requested.

Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

Gaby Longsworth

Gaby L. Longsworth
Agent for Applicants
Registration No. 47,756

Date: July 24, 2001

1100 New York Avenue, N.W.
Suite 600
Washington, D.C. 20005-3934
(202) 371-2600

Version with markings to show changes made

In the specification:

Fourth full paragraph on page 2, lines 24-26:

Figures [1A-C] 1A-B show the nucleotide sequence (SEQ ID NO:1) and the deduced amino acid sequence (SEQ ID NO:2) of PDEF. Regions of conservation to the pointed and Ets domains are indicated by single and double underline, respectively.

Second full paragraph on page 3, lines 16-22:

[Figure 4 shows] Figures 4A-C show the tissue distribution of PDEF expression in different human fetal (Figure 4A) and adult tissues (Figures 4B-C) by Northern hybridization. The blots were sequentially probed with PDEF (upper panel), ESE-1 (middle panel) and GADPH cDNA probes (lower panel) under stringent conditions using poly(A)+ mRNA from the indicated tissues (See Example 3 of present invention). A skilled artisan would readily associate the intensity and location of the bands with respect to the blot as indicative of both the abundance and size of the PDEF mRNA within each tissue.

Third full paragraph on page 3, lines 24-28:

[Figure 5 shows] Figures 5A-B show the tissue distribution of PDEF expression within poly(A)+ mRNA from human fetal and adult tissues by Dot Blot Hybridization. The blot was probed with PDEF under the conditions described in Example 3 of present invention. A skilled artisan would readily associate the intensity of the dots as indicative of both the abundance of the PDEF mRNA within each tissue.

Fourth full paragraph on page 3, lines 30-33:

[Figure 6 shows] Figures 6A-D show [in situ] *in situ* hybridization studies. Paired brightfield (A, C) and corresponding polarized fluorescence (B, D) photomicrographs. Intense labeling of prostate epithelium in normal lung is seen with antisense probe to PDEF mRNA (A, B). No labeling is seen with control sense probe (C, D).

In the claims:

Please rewrite the claims as follows:

24. (Twice Amended) An isolated polynucleotide comprising a first nucleic acid at least 90% identical to a reference nucleic acid selected from the group consisting of:

(a) a nucleic acid encoding amino acids 142 to 211 of SEQ ID NO:2;
(b) a nucleic acid encoding amino acids 248 to 331 of SEQ ID NO:2;
(c) a nucleic acid encoding amino acids 2 to 335 of SEQ ID NO:2;
(d) a nucleic acid encoding amino acids 1 to 335 of SEQ ID NO:2; and

(e) a nucleic acid encoding the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 203072;
wherein said first nucleic acid encodes a polypeptide which [generates an antibody that] binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2.

51. (Twice amended) An isolated polynucleotide comprising a nucleic acid encoding a first amino acid sequence at least 90% identical to a reference amino acid sequence selected from the group consisting of:

(a) amino acids 142 to 211 of SEQ ID NO:2;

(b) amino acids 248 to 331 of SEQ ID NO:2;

(c) amino acids 2 to 335 of SEQ ID NO:2;

(d) amino acids 1 to 335 of SEQ ID NO:2; and

(e) the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 203072;
wherein said nucleic acid encodes a polypeptide which [generates an antibody that] binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2.

105. (Three times amended) An isolated polynucleotide comprising a nucleic acid at least 95% identical to a nucleic acid encoding at least 70 contiguous amino acids of SEQ ID NO:2;
wherein said nucleic acid encodes a polypeptide which [generates an antibody that] binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2.

107. (Three times amended) An isolated polynucleotide comprising a nucleic acid at least 95% identical to a nucleic acid encoding at least 80 contiguous amino acids of SEQ ID NO:2;
wherein said nucleic acid encodes a polypeptide which [generates an antibody that] binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2.

109. (Twice amended) An isolated polynucleotide comprising a nucleic acid at least 95% identical to a nucleic acid encoding at least 100 contiguous amino acids of SEQ ID NO:2;
wherein said nucleic acid encodes a polypeptide which [generates an antibody that] binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2.

111. (Twice amended) An isolated polynucleotide comprising a nucleic acid at least 95% identical to a nucleic acid encoding at least 150 contiguous amino acids of SEQ ID NO:2;
wherein said nucleic acid encodes a polypeptide which [generates an antibody that] binds with specificity to antibodies having specificity for a polypeptide consisting of amino acids 1 to 335 of SEQ ID NO:2.

128. (Three times amended) An isolated polynucleotide comprising a nucleic acid encoding an amino acid sequence from position m to position n of SEQ ID NO:2, wherein m is an integer from 2 to [276] 236, n is an integer from [141] 243 to 335; and m is less than n.

137. (Three times amended) A polynucleotide comprising a nucleic acid fused in frame to a nucleotide sequence heterologous to SEQ ID NO:1, wherein said heterologous nucleotide sequence encodes a heterologous polypeptide, and wherein said nucleic acid is selected from the group consisting of:

(a) a nucleic acid encoding amino acids 279 to 287 of SEQ ID NO:2;
(b) a nucleic acid encoding amino acids 292 to 300 of SEQ ID NO:2; and
(c) a nucleic acid encoding amino acids 317 to 325 of SEQ ID NO:2.

154. (Once Amended) A method of producing [a] the polypeptide encoded by the polynucleotide of claim 149, comprising culturing [the] a host cell [of claim 153] comprising said polynucleotide under conditions such that [a] said polypeptide is expressed, and recovering said polypeptide.

New claim 156 was added.

EXPRESSION OF PROTEINS IN *ESCHERICHIA COLI*

Overview of Protein Expression in *E. coli*

The study of *Escherichia coli* during the 1960s and 1970s made it the best understood organism in nature (Chapter 1). Today's recombinant DNA technology is a direct extension of the genetic and biochemical analyses carried out at that time. Even before the advent of molecular cloning, genetically altered *E. coli* strains were used to produce quantities of proteins of scientific interest. When cloning techniques became available, most cloning vectors utilized *E. coli* as their host organism. Thus, it is not surprising that the first attempts to express large quantities of proteins encoded by cloned genes were carried out in *E. coli.*

E. coli has two characteristics that make it ideally suited as an expression system for many kinds of proteins: it is easy to manipulate and it grows quickly in inexpensive media. These characteristics, coupled with more than 10 years' experience with expression of foreign genes, have established *E. coli* as the leading host organism for most scientific applications of protein expression.

Despite a growing literature describing successful protein expression from cloned genes, each new gene still presents its own unique expression problems. No one, and certainly no laboratory manual, can provide a set of methods that will guarantee successful production of every protein in a useful form. Nevertheless, the vast body of accumulated knowledge has led to a general approach that often helps to solve specific expression problems. This unit introduces general considerations and strategies, while subsequent units (*16.2-16.7*) describe procedures that can be applied to specific expression problems.

GENERAL STRATEGY FOR GENE EXPRESSION IN *E. COLI*

The basic approach used to express all foreign genes in *E. coli* begins with insertion of the gene into an expression vector, usually a plasmid. This vector generally contains several elements: (1) sequences encoding a selectable marker that assure maintenance of the vector in the cell; (2) a controllable transcriptional promoter (e.g., *lac, trp,* or tac) which, upon induction, can produce large amounts of mRNA from the cloned gene; (3) translational control sequences, such as an appropriately positioned ribosome-binding site and initiator ATG; and (4) a polylinker to simplify the insertion of the gene in the correct orientation within the vector. Once constructed, the expression vector containing the gene to be expressed is introduced into an appropriate *E. coli* strain by transformation (*UNIT 1.8*).

SPECIFIC EXPRESSION SCENARIOS

Although this general approach—insertion of the gene of interest into an expression vector followed by transformation in *E. coli*—is common to all expression systems, specific procedures differ greatly. When choosing a procedure, it is helpful to consider the final application of the expressed protein, as this often dictates which expression strategy to use (*UNIT 16.4A*).

Antigen Production

If the goal is to use the expressed protein as an antigen to make antibodies, several approaches are available to make protein reliably and to allow for rapid purification of the antigen. The two best approaches are synthesis of fusion proteins with specific "tag" sequences that can be retrieved by affinity chromatography (*UNITS 16.5, 16.6, 16.7 & 16.8;* see also *UNIT 10.11B*) and synthesis of the native protein, or a fragment of it, under conditions that cause it to precipitate into insoluble inclusion bodies (*UNITS 16.4A & 16.6*). These inclusion bodies can be purified sufficiently by differential centrifugation so preparative denaturing polyacrylamide gel electrophoresis (*UNIT 10.2*) will yield an isolated band that can be cut out and crushed, or electroeluted (*UNIT 10.5*), to provide antigenic material for injection into an animal.

Biochemical or Cell Biology Studies

If the goal is to use the expressed protein as a reagent in a series of biochemical or cell biology experiments, other considerations are relevant. In this case, the authenticity of the protein's function (e.g., high-specific-activity enzyme, binding protein, or growth factor) is

very important, while the ease of preparing the protein matters less. For this application, it is possible to express the protein as a fusion protein containing a specific protease-sensitive cleavage site so the N-terminal peptide tail can be removed easily, leaving only the native amino acid sequence (*UNITS 16.4, 16.6, 16.7 & 16.8*). Alternatively, direct expression vectors of the type described in *UNITS 16.2 & 16.3* may be used to produce the authentic primary sequence. When expressed, the protein may be soluble and active, as is the case with many intracellular enzymes. If it is insoluble, as is the case for many secreted growth factors when they are made cytoplasmically in *E. coli*, it may be necessary to isolate inclusion bodies, solubilize the protein using denaturing agents, and refold the protein. Refolding is usually not too difficult when the protein is of moderate size (Marston and Hartley, 1990). Whether the protein is expressed in a soluble form or whether it requires refolding, its integrity can usually be checked by specific enzyme assays or by bioassays.

Structural Studies

If the goal is to do structural studies of the expressed protein, the greatest constraints are imposed on the expression system. Because it is nearly impossible to show that a protein of unknown structure has been precisely refolded after denaturing, the protein must generally be made in a soluble form so its purification does not require a denaturation/renaturation step. Usually, the soluble form of the protein—either intracellular or secreted—must be made in strains and by induction protocols that minimize proteolytic degradation.

Soluble expression of most eukaryotic proteins is best achieved with systems that allow induction of synthesis without changing the temperature; for example, by inducing transcription from the *trp* (Edman et al., 1981; de Boer et al., 1983) or tac (de Boer et al., 1983) promoters. Maximum accumulation of soluble product is best achieved by testing expression in several strains and at several temperatures, and picking the combination that works best. This is an active area of research at present (Schein, 1989); the rules are not yet understood, so little more than trial and error can be recommended.

TROUBLESHOOTING GENE EXPRESSION

Once an expression strategy has been chosen and the gene is introduced into an appropriate expression vector, several strains of *E. coli* should be transformed with the vector and protein production should be monitored. Ideally, the protein of interest will be produced in an active form and in sufficient amounts to allow its isolation. Often, however, the protein will be made either in very small amounts or in an insoluble form, or both. If this happens, there are various approaches that may correct the problem.

If not enough protein is produced:

1. Reconstruct the 5′-end of the gene, maximizing its A+T content while preserving the protein sequence it encodes. This may reduce secondary structure within the mRNA (DeLamarter et al., 1985), or it may alter an as yet undefined parameter of the reaction. Regardless of the underlying cause, this procedure usually increases translation efficiency.

2. Determine if a transcriptional terminator is present. If the vector does not have a transcriptional terminator downstream from the site at which the gene is inserted, put one in. This often aids expression, probably by increasing mRNA stability and by decreasing nucleotide drain on the cell.

3. Examine the sequence of the cloned gene for codons used infrequently in *E. coli* genes. These so-called rare codons are usually not a rate-limiting problem, but if four or more happen to occur contiguously, they can reduce expression significantly (Robinson et al., 1984), perhaps by causing ribosomes to pause. Ribosomal pausing can uncouple transcription from translation, leading to premature termination of the message. Even if transcription proceeds normally, the mRNA 3′ to the stalled ribosomes can be exposed to degradation by host ribonucleases, reducing its stability. Thus, if stretches of rare codons are found, they should be altered to codons more favorable to high expression in *E. coli*.

If enough protein is produced, but it is insoluble when the application requires it to be active and soluble:

1. Vary the growth temperature. As mentioned above, many proteins are more soluble at lower than at higher temperatures (Schein and Noteborn, 1988). On the other hand, some enzymes have a higher specific activity when made at temperatures >37°C (J. McCoy and P. Schendel, unpub. observ.). *E. coli* can synthesize proteins at temperatures ranging from 10° to 43°C, so trying expression at different temperatures is often worthwhile.

2. Change fermentation conditions. Many proteins contain metals as structural and catalytic cofactors. If the protein is being made faster than metals can be transported into the cell, the apoprotein without its metal cofactor will accumulate. This apoprotein will not fold correctly and will likely be insoluble. At the very least, the average specific activity of the expressed protein will be lower than expected. Different media and metal supplements can be tested and the best combination used. Clearly, if there is information about the metal content of the protein, these supplements can be designed more rationally. If no information is available, a more random approach must be tried.

3. Alter the rate of expression by using low-copy-number plasmids. This can be done by using the pACYC family (Chang and Cohen, 1978) or using single-copy chromosomal inserts of the cloned gene into a suitable target gene (Hamilton et al., 1989). Such reductions in gene dosage often reduce the final yield of protein, but the slower kinetics of synthesis they afford can sometimes result in production of soluble proteins.

To restate the obvious, protein expression is an inexact science at present. However, most proteins can be made in *E. coli* in a form that is useful for a variety of functions. The procedures employed are relatively quick and uncomplicated, and the rewards for success are great.

Literature Cited

Chang, A.C.Y. and Cohen, S.N. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* 134:1141-1156.

de Boer, H.A., Comstock, L.J., and Vasser, M. 1983. The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. *Proc. Nat. Acad. Sci. U.S.A.* 80:21-25.

DeLamarter, J.F., Mermod, J.J., Liang, C.M., Eliason, J.F., and Thatcher, D.R. 1985. Recombinant murine GM-CSF from *E. coli* has biological activity and is neutralized by a specific antiserum. *EMBO J.* 4:2575-2581.

Edman, J.C., Hallewell, R.A., Valenzuela, P., Goodman, H.M., and Rutter, W.J. 1981. Synthesis of Hepatitis B surface and core antigens in *E. coli*. *Nature* 291:503-506.

Hamilton, C.M. Aldea, M., Washburn, B.K., Babitzke, P., and Kushner, S.R. 1989. New methods for generating deletions and gene replacements in *Escherichia coli*. *J. Bacteriol.* 171:4617-4622.

Marston, F.A.O. and Hartley, D.L. 1990. Solubilization of protein aggregates. *Methods Enzymol.* 182:264-276.

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M., and Humphreys, G. 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucl. Acids Res.* 12 6663-6671.

Schein, C.H. 1989. Production of soluble recombinant proteins in bacteria. *Bio/Technology* 7:1141-1148.

Schein, C.H. and Noteborn, M.H.M. 1988. Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperatures. *Bio/Technology* 6:291-294.

Contributed by Paul F. Schendel
Genetics Institute
Cambridge, Massachusetts

EXPRESSION OF PROTEINS IN MAMMALIAN CELLS

Overview of Protein Expression in Mammalian Cells

The techniques of gene isolation, modification, and transfer into appropriate host cells have provided a powerful means to study gene expression and to evaluate protein structure and function. These techniques make it possible to produce large amounts of proteins that previously could be isolated in only minute quantities and allow the generation of proteins with specific, designed alterations.

As described elsewhere in this manual (Chapter 9), mammalian cells are often used as hosts for expression of genes obtained from higher eukaryotes because the signals for synthesis, processing, and secretion of these proteins are usually recognized. Mammalian expression of foreign genes has been used for many purposes including: (1) confirmation that genes isolated by different approaches can direct the synthesis of a desired protein, (2) evaluation of the effect of specific mutations introduced into genes, (3) direct isolation of genes based on screening or selecting recipient cells for production of a particular protein, (4) production of large amounts of proteins that normally are available in only limited quantity, and (5) analysis of physiologic consequences of expression of specific proteins.

The units that follow in Sections III and IV describe three vector systems or strategies for introducing foreign genes into mammalian cells (additional transfection methods can be found in *UNITS 9.1-9.4*). The first method relies upon COS cells for rapid, transient expression of protein from specific vectors (*UNIT 16.13*). The second method relies upon Chinese hamster ovary (CHO) cells; in this procedure, a targeted gene is cotransfected with a selectable marker, becomes stably integrated into the host cell chromosome, and is subsequently amplified (*UNIT 16.14*). The third system relies upon vaccinia viral vectors in a transient expression system (*UNITS 16.15-16.19*). The systems described in these units differ in the ways in which DNA is introduced into the cell, in the particular vectors used with each system, and in their suitability for particular cell types.

The criteria for choosing a certain system include these considerations: whether DNA can be introduced directly by transfection methods or needs to be introduced by viral-mediated transfer, the identity of the control elements that can direct efficient mRNA expression and protein synthesis, and whether a particular host cell is appropriate for expression of the gene of interest. If it is necessary to produce a large amount of protein for a long period of time, the CHO system should be utilized. When transient expression is appropriate, the choice of which system to use depends upon the particular experiment. When a high transfection efficiency is necessary, the vaccinia system is appropriate because every cell can be infected with the virus and gene of interest; however, this system suffers from the disadvantage that the cells die within one to two days. If a lower transfection efficiency is sufficient and if it is desirable that the cells continue to grow for several days, COS cells should be used.

VIRAL-MEDIATED GENE TRANSFER

Many viruses that infect mammalian cells have evolved mechanisms to usurp the protein synthesis machinery of the host to produce their viral proteins. The ability to engineer the genetic material of these viruses has made it possible to place desired coding regions under the control of the viral expression elements and to produce infectious virus particles that direct high levels of foreign gene expression. Viral-mediated gene transfer provides a convenient, efficient means to introduce foreign DNA into the majority of recipient cells. In addition, for many viruses, viral replication yields multiple copies of template DNA that can serve to increase the total amount of transcript made by the foreign gene.

Because some viruses can infect a range of cell types derived from different species, viral-mediated gene transfer often allows the convenient introduction of foreign genes into a variety of different cells. Representative

Contributed by Randal J. Kaufman

Current Protocols in Molecular Biology (1997) 16.12.1-16.12.6

Table 16.12.1 Expression Levels and Uses for Different Mammalian Cell Expression Systems[a]

Cell line	Expression method	Typical expression level (μg/ml)	Primary use
Monkey cells			
CV1	SV40 virus infection	1-10	Expression of wild-type and mutant proteins
COS	Transient DNA/ DEAE-dextran transfection	1	Cloning by expression in mammalian cells; rapid characterization of cDNA clones; expression of mutant proteins
CV1	Transient DNA/ DEAE-dextran transfection	0.05	
Murine fibroblasts			
C127	BPV stable transformant	1-5	High-level constitutive protein expression
3T3	Retrovirus infection	0.1-0.5	Gene transfer into animals; expression in different cell types
Other cells			
CHO(DHFR$^-$)	Stable DHFR$^+$ transformant	0.01-0.05	
	Amplified MTX^r	10	High-level constitutive protein expression
Primate	Vaccinia virus infection	1	Production of vaccines; expression of toxic proteins
	EBV vector	n.a.	Cloning by expression

[a]Abbreviations: BPV, bovine papilloma virus; EBV, Epstein-Barr virus.

expression levels obtained from SV40 recombinant viruses, retroviruses, and vaccinia viruses are shown in Table 16.12.1 in comparison to other expression strategies. A more detailed review of the different eukaryotic viral vectors can be found in Muzyczka (1989).

Most viral expression systems have certain common limitations. The first is size of the inserted sequence. If the sequence is too large, it may not be packaged properly into the viral genome (maximal insert sizes for SV40 and retroviruses are 2.5 kb and 6 kb, respectively) and/or may be subjected to rearrangement upon propagation of the viral stock. The second limitation is the cytopathic effect of some viruses on the host cell, which limits expression to a relatively short period of time. Third, the variability in gene expression depends on multiple parameters. The reasons for this are not completely clear but depend upon proper translation, processing, and modification of the resulting protein; thus, there is a large degree of variability in the success with any particular DNA insert.

UNITS 16.15-16.19 address one viral vector system, vaccinia virus, which has demonstrated success. Vaccinia virus is most useful for the production of proteins (such as regulatory factors) that are potentially toxic to the cell.

TRANSIENT EXPRESSION

The efficiency of expression from transient transfection depends on the number of cells that take up the transfected DNA, the gene copy number, and the expression level per gene. Most methods of DNA transfer allow 5% to 50% of the cells in the population to acquire DNA and express it transiently over a period of several days to several weeks. Eventually, however, because cells containing the foreign DNA grow more slowly, they are lost from the population. Although this gradual decrease in the amount of expression occurs, transient expression offers a convenient means to compare

expression from different vectors and to verify that any given expression plasmid is functional before initiating the more laborious procedure of isolating and characterizing stably transfected cell lines. Because they do not require isolation of clones of cells with the vector integrated at different sites, transient expression experiments are not subject to the differences such position effects can have on expression levels. In addition, transient expression experiments preclude the possibility of selecting cells that harbor mutations either in the transfected DNA or elsewhere in the host chromosome (which may influence results).

Transient DNA transfection is most frequently used to: (1) verify the identity of cloned genes based on their ability to express a particular activity, (2) rapidly study the effect of engineered mutations on either gene activity or protein function, and (3) isolate genes from cDNA libraries constructed in mammalian expression vectors based on their ability to express a particular activity in cells. The limitations of transient expression are that it is difficult to scale up for production of large quantities of protein (>1 mg), that it is difficult to study the consequences of gene expression only in the portion of the total population that has been transfected, and that the high copy number is eventually lethal; this lethality may significantly affect results.

UNIT 16.13 describes procedures and vectors used for transient expression in COS cells. This cell line is most frequently used for transient expression and is derived from African green monkey kidney cells by tranformation with an origin-defective simian virus 40 (SV40). COS cells express high levels of the SV40 large tumor (T) antigen which is required to initiate viral DNA replication at the origin of SV40. T antigen–mediated replication can amplify the copy number of plasmids containing the SV40 origin of replication to >100,000 per cell, which results in high expression levels from the transfected DNA.

STABLE DNA TRANSFECTION

If a selection procedure is applied after DNA transfection, it is possible to isolate cells that have stably integrated the foreign DNA into their genome (see *UNIT 9.5* for an additional discussion). Different cell lines exhibit different frequencies of stable transformation and vary in their capacity to incorporate foreign DNA. In most cases, the limiting factor for obtaining stable transformants is the frequency of DNA integration, not the frequency of DNA uptake. Cells selected for by incorporation and expression of one genetic marker will frequently incorporate a second gene provided on an independent plasmid during transfection; this ability to incorporate two separate plasmids into the chromosome has been termed cotransformation. During this process, transfected DNA molecules usually become ligated inside the cell and subsequently integrate as a unit via nonhomologous recombination into host chromosomal DNA.

Different cell lines and transfection methods yield varying frequencies of cotransformation. For example, the frequency of stable cotransformation in CHO cells is lower than that observed in mouse L cells, possibly because mouse L cells are able to incorporate more DNA into their genome than CHO cells. Cotransformation using calcium phosphate–mediated DNA transfection is very efficient, whereas cotransformation by fusion of bacterial protoplasts (containing two independent plasmids) with mammalian cells is very rare. When cotransformation is inefficient, it is desirable to engineer both the selectable marker and the gene of interest into the same plasmid. A number of vectors that facilitate this procedure have been constructed (Kaufman, 1990a).

Stable transformants are usually selected by their ability to confer resistance to cytotoxic drugs. Such resistance can be recessive or dominant. Genes conferring dominant drug resistance can be used independently of the host cell line. Frequently, selectable marker genes are derived from bacterial genes for which there is no mammalian counterpart. For example, the Tn*5* neomycin phosphotransferase gene encoding resistance to the antibiotic G418, or the *Escherichia coli* hygromycin phosphotransferase gene encoding resistance to hygromycin have been engineered to be expressed and selected in mammalian cells.

Genes conferring recessive drug resistance require a particular host that is deficient in the activity being selected. Many recessive genetic selectable markers encode enzymes involved in the purine and pyrimidine biosynthetic pathways. When de novo biosynthesis for purines or pyrimidines is inhibited, cells can utilize purine or pyrimidine salvage pathways, providing the salvage enzymes (i.e., thymidine kinase, hypoxanthine-guanine phosphoribosyltransferase, adenine phosphoribosyltransferase, or adenosine kinase) necessary for the conversion of nucleoside precursors to the corresponding nucleotides are present (Fig 16.12.1). These salvage enzymes are not required for cell

growth when de novo purine or pyrimidine biosynthesis is functional; thus, cells deficient in a particular salvage pathway are viable under normal growth conditions. However, when drugs that inhibit de novo biosynthesis (such as methotrexate) are added, the cells die. Cells that acquire the capability to express the deficient activity (i.e., the missing salvage enzyme) via gene tranfer can be selected for under these conditions. In a complementary manner, in cells that are defective in de novo biosynthetic pathways and have functional salvage pathways, it is possible to select for expression of the defective gene (e.g., dihydrofolate reducatase or aspartate transcarbamylase) in the de novo pathway by removal of nucleosides from the growth medium.

AMPLIFICATION OF TRANSFECTED DNA

Frequently, it is desirable to increase expression by selecting for increased copy number of the transfected DNA within the host chromosome. The ability to coamplify transfected DNA has permitted a 100- to 1000-fold increase in the expression of the proteins encoded by transfected DNA. Although there are over twenty selectable and amplifiable genes that have been described (Kaufman, 1990a), the most experience and success has occurred when methotrexate selection has been used for amplification of transfected dihydrofolate reductase genes. *UNIT 16.14* describes the use of dihydrofolate reductase–deficient CHO cells to obtain high-level expression of heterologous

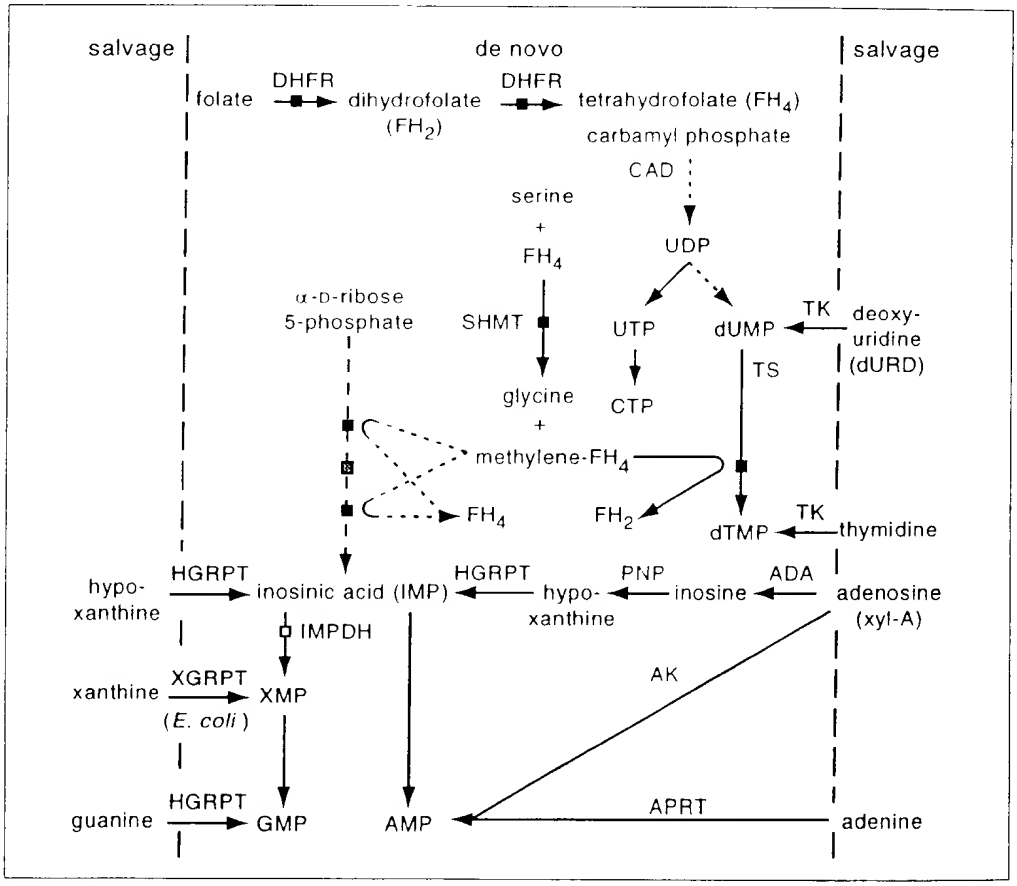


Fig. 16.12.1 Purine and pyrimidine biosynthetic pathways. The solid arrows indicate single reactions. Dashed arrows indicate multiple reactions. Solid squares indicate reactions inhibited by methotrexate. The hatched square indicates the primary reaction inhibited by azaserine. The open square indicates the reaction inhibited by mycophenolic acid. Abbreviations used for enzymes involved in de novo biosynthetic pathways: **DHFR**, dihydrofolate reductase; **CAD**, carbamoyl-phosphate synthetase, aspartate transcarbamoylase, and dihydroorotase; **SHMT**, serine hydroxymethyl transferase; **TS**, thymidylate synthetase; **IMPDH**, inosine-monophosphate dehydrogenase. Abbreviations used for enzymes involved in salvage pathways: **TK**, thymidine kinase; **ADA**, adenosine deaminase; **PNP**, purine nucleoside kinase; **AK**, adenosine kinase; **APRT**, adenosine phosphoribosyltransferae; **HGPRT**, hypoxanthine-guanine phosphoribosyltransferase; **XGPT**, *E. coli* xanthine-guanine phosphoribosyltransferase. Other abbreviations: **FH**, tetrahydrofolate; **xyl A**, 9-β-D-xylofuranosyl adenine. This figure has been adapted from Kaufman (1987).

genes through coamplification by selection for methotrexate resistance.

EXPRESSION VECTORS

Although a wide variety of expression vectors have been described over the past 5 years, it is difficult to compare results from different vectors used in different laboratories with different inserts. Most mammalian cell expression vectors are designed to accomodate cDNAs rather than large genomic fragments because the small size of cDNA clones makes them more convenient to manipulate. Today most useful vectors contain multiple elements including: (1) an SV40 origin of replication for amplification to high copy number in COS monkey cells, (2) an efficient promoter element for high-level transcription initiation, (3) mRNA processing signals such as mRNA cleavage and polyadenylation sequences, and frequently intervening sequences as well, (4) polylinkers containing multiple restriction endonucleases sites for insertion of foreign DNA, (5) selectable markers that can be used to select cells that have stably integrated the plasmid DNA, and (6) plasmid backbone sequences to permit propagation in bacterial cells.

In addition to the previously mentioned properties, most vectors also contain an inducible expression system that is regulated by an external stimulus. Sequences from a number of promoters that are required for induced transcription have been identified and engineered into expression vectors to obtain inducible expression. Several useful inducible vectors have been based on induction by β-interferon, heat-shock, heavy metal ions, and steroids such as glucocorticoids (Kaufman, 1990b). If the effect of an expressed protein on a particular cellular process is being studied, it is important to determine that the inducing stimulus does not interfere with that cellular process. It is also important to know the factor of induction (i.e., the difference between the basal and induced level of expressed protein) as well as the maximal achievable expression level. In many cases, the factor of induction may be large but the maximal level of expression very low compared to expression from a strong constitutive promoter.

CHOICE OF EXPRESSION SYSTEM

In evaluating which approach to take in expressing a gene, it is most important to consider the goals of the expression work. If expression is required to demonstrate that a clone has some functional activity or to characterize this activity, then transient expression in COS cells is often the most convenient approach. If a large quantity of protein (>1 mg) is required, then stable coamplification in CHO cells is generally the most desirable approach. If the gene is potentially cytotoxic, high-level expression may be approached through vaccinia virus vectors or inducible promoter-vector systems. If there is a particular requirement for the host to produce the protein properly then that requirement will dictate the choice of the host. It is unusual that proteins do require host-specific posttranslational modifications. However, if this does occur, it is usually desirable to study expression of the gene in a variety of host cells, and a retrovirus vector would be the system of choice.

TROUBLESHOOTING

If protein expression from the heterologous gene cannot be detected, it is important to examine the vector system in detail. In this sequence, each point should be satisfactorily addressed before proceeding to the next step.

1. Confirm the expected structure of the vector using restriction mapping (*UNITS 3.1 & 3.2*) and, if necessary, DNA sequencing (*UNITS 7.4 & 7.5*).

2. Determine transfection efficiency by including a positive control—e.g., the same vector with another insert.

3. Ensure that the RNA is of the expected size and amount compared to an appropriate control by preparing RNA (*UNITS 4.1, 4.2, & 9.8*) and analyzing it by northern hybridization (*UNIT 4.9*).

4. Use a completely different expression vector or system (*UNITS 16.13-16.18*) if the RNA transcript of the correct size cannot be detected in the transfected cells, as it is always possible that some unforeseen situation may result in aberrant splicing (Wise et al., 1989).

5. Determine if the coding region may contain a point mutation or other lesion that keeps it from encoding a full-length protein by carrying out in vitro translation to produce protein (*UNIT 10.17*) using mRNA isolated from transfected cells and using RNA transcribed by in vitro transcription (i.e., SP6; *UNIT 3.8*) of a vector containing the cDNA insert.

LITERATURE CITED

Kaufman, R.J. 1987. High level production of proteins in mammalian cells. *In* Genetic Engineering: Principles and Methods (J.K. Setlow, ed.) pp. 155-198. Plenum, New York.

Kaufman, R.J. 1990a. Selection and coamplification of heterologous genes in mammalian cells. *Methods Enzymol.* 185:537-566.

Kaufman, R.J. 1990b. Overview of vectors used for expression in mammalian cells. *Methods Enzymol.* 185:487-511.

Muzyczka, N. (ed.) 1989. Eukaryotic Viral Expression Vectors: Current Topics in Microbiology and Immunology. Springer-Verlag, Berlin.

Wise, R.J., Orkin, S.H., and Collins, T. 1989. Aberrant expression of platelet-derived growth factor A-chain cDNAs due to cryptic splicing of RNA transcripts in COS-1 cells. *Nucl. Acid Res.* 17:6591-6601.

Contributed by Randal J. Kaufman
University of Michigan
Ann Arbor, Michigan

Transient Expression of Proteins Using COS Cells

BASIC PROTOCOL

This unit describes the use of COS cells to efficiently produce a desired protein in a short period of time. These cells express high levels of the SV40 large tumor (T) antigen, which is necessary to initiate viral DNA replication at the SV40 origin. Three factors contribute to make COS cell expression systems appropriate for the high-level, short-term expression of proteins: (1) the high copy number achieved by SV40 origin–containing plasmids in COS cells 48 hr posttransfection, (2) the availability of good COS cell expression/shuttle vectors, and (3) the availability of simple methods for the efficient transfection of COS cells. Each COS cell transfected with DNA encoding a cell-surface antigen (in the appropriate vector) or cytoplasmic protein will express several thousand to several hundred thousand copies of the protein 72 hr posttransfection. If the transfected DNA encodes a secreted protein, up to 10 µg of protein can be recovered from the supernatant of the transfected COS cells 1 week posttransfection. COS cell transient expression systems have also been used to screen cDNA libraries, to isolate cDNAs encoding cell-surface proteins, secreted proteins, and DNA binding proteins, and to test protein expression vectors rapidly prior to the preparation of stable cell lines (*UNIT 9.5*).

This transfection protocol is a modification of that presented in *UNIT 9.2* and gives conditions for optimal transfection of COS cells (see *UNIT 9.9* for additional details). The main difference between this procedure and that in *UNIT 9.2* is the composition of the DEAE-dextran/chloroquine solution, which is prepared here in PBS, not TBS, and contains chloroquine to prevent the acidification of endosomes presumed to carry the DEAE-dextran/DNA into the cell. (This acidification results in acid hydrolysis of the DNA, giving rise to mutations and destruction of the DNA.) With this protocol, 40% to 70% of the cells can be routinely transfected.

Materials

Appropriate vector (see commentary)
COS-7 cells to be transfected (see critical parameters; ATCC #CRL1651)
Dulbeccos minimum essential medium with 10% calf serum (DMEM-10 CS; *UNIT 9.4*)
DMEM with 10% NuSerum (Collaborative Research #55000) (DMEM-10 NS; *UNIT 9.4*), 37°C
Phosphate-buffered saline (PBS; *APPENDIX 2*)
DEAE-dextran/chloroquine solution: PBS containing 10 mg/ml DEAE-dextran (Sigma #D9885) + 2.5 mM chloroquine (Sigma #C6628)
10% dimethyl sulfoxide (DMSO; Sigma #D5879) in PBS
0.5 mM EDTA in PBS

100-mm tissue culture dishes
Humidified 37°C, 6% CO_2 incubator
Phase-contrast microscope
Sorvall RT-6000B rotor (or equivalent)

Additional reagents and equipment for subcloning DNA (*UNIT 3.16*), preparing miniprep DNA (*UNIT 1.6*), and purifying DNA by CsCl/ethidium bromide equilibrium centrifugation (*UNITS 1.7 & 9.1*)

1. Subclone the gene of interest into the appropriate vector to obtain the desired recombinant DNA. Purify the recombinant DNA by a miniprep procedure (5-ml culture) or by CsCl/ethidium bromide centrifugation.

Contributed by Alejandro Aruffo
Current Protocols in Molecular Biology (1997) 16.13.1-16.13.7

2. Seed COS-7 cells in DMEM-10 CS at ~20% confluence per 100-mm dish the day prior to transfection (so they will be ~50% confluent the next day). Grow cells overnight in a CO_2 incubator (6% CO_2) at 37°C to ~50% confluence.

A confluent dish of COS-7 cells (~10^6 cells) is usually split 1:5 on the day prior to transfection to give 2 × 10^5 cells/100-mm dish in 10 ml of DMEM-10 CS.

3. Just before use (for each 100-mm dish of COS cells to be transfected), thoroughly mix 5 ml of 37°C DMEM-10 NS with 0.2 ml of DEAE-dextran/chloroquine solution. Add 5 to 10 µg recombinant DNA and mix.

It is important that the DEAE-dextran be well mixed with the media before adding the DNA—otherwise, the DNA, a negatively charged polymer, will form large precipitates with DEAE-dextran, a positively charged polymer. These large precipitates cannot be taken up by the cell, resulting in a reduced transfection efficiency. When larger dishes are used, the amount of medium/DEAE/DNA should be sufficient to easily cover the cells and should include 400 µg/ml DEAE-dextran, 100 µM chloroquine, and 1 to 2 µg/ml DNA in DMEM-10 NS.

Either CsCl-purified (UNIT 1.7) or miniprep (UNIT 1.6) plasmid DNA can be used for the transfections. If miniprep DNA is used, use ⅓ of the miniprep per transfection. DNA can also be prepared as described in the second support protocol of UNIT 9.1.

4. Aspirate medium from COS cells and for each 100-mm dish, add DMEM-10 NS/DEAE-dextran/DNA prepared in step 3. Incubate cells 3 to 4 hr in a CO_2 incubator at 37°C. Observe the cells using a phase-contrast microscope.

The DEAE-dextran will cause cells to retract and become vacuolated. Efficiency of transfection increases with longer incubation periods; on the other hand, so does cell death. The 3- to 4-hr incubation suggested here is a good starting point. However, several time points should be tried to optimize transfection of the particular population of cells used.

5. Aspirate DMEM/DEAE-dextran/DNA and add 5 ml of 10% DMSO (prepared in PBS). Incubate cells 2 min at room temperature.

The DMSO shock results in increased transfection efficiencies. Without this step, transfection efficiencies might be lower by a factor of two or more.

6. Aspirate DMSO and add 10 ml DMEM-10 CS. Grow cells overnight (12 to 20 hr) in a CO_2 incubator at 37°C.

7. Passage (split and replate) each 100-mm dish of transfected COS cells into two new 100-mm dishes. Grow cells in a CO_2 incubator at 37°C as described in step 8a or 8b.

After transfection, the COS cells will look unhealthy. Passaging them the day after transfection facilitates recovery, resulting in better levels of protein expression. In addition, DEAE-dextran treatment makes the cells sticky, and passaging the cells the morning after transfection restores their adhesion characteristics so that they may be once again lifted by a gentle treatment with PBS and EDTA (see step 8b).

8a. When expressing secreted proteins, add 5 ml DMEM-10 CS 96 hr (4 days) after completing step 7 and incubate 4 days. Harvest the medium, remove dead cells and debris by centrifuging 10 min in a Sorvall RT-6000B rotor at ~2000 rpm (~1000 × *g*), room temperature, and save the supernatant (see anticipated results). Detect secreted proteins by metabolic labeling (*UNIT 10.18*) and immunoprecipitation (*UNIT 10.16*), immunoaffinity chromatography (*UNIT 10.11*), radioimmunoassay (*UNIT 11.16*), western blotting (*UNIT 10.8*), or bioassay (*UNIT 9.5*).

Do not aspirate the old medium prior to addition of 5 ml DMEM-10 CS because this medium contains the secreted protein. Addition of extra medium 96 hr posttransfection results in better yield of expressed protein; however, it also increases the level of total

protein (since the medium contains 10% serum), which could complicate protein purification. To eliminate this problem, COS cells can be placed in serum-free medium 10 to 12 hr after they have been replated although (in our hands) this results in a 10-fold lower yield of expressed protein than in the presence of serum. Thus, unless it is absolutely necessary to remove additional contaminating protein, serum should be present in the medium even at reduced levels (1% to 2%).

8b. When expressing cell-surface or intracellular proteins, aspirate medium from cells 72 hr (3 days) after transfection in step 6. Add 5 ml PBS, swirl, and aspirate PBS. Add 5 ml of 0.5 mM EDTA in PBS and incubate 15 min in a CO_2 incubator at 37°C. Lift cells from the dish by gently dislodging them with a Pasteur pipet. Stain cell-surface proteins with the appropriate fluorescent antibody and detect by microscopy or flow cytometry (Yokoyama, 1991).

Transfected COS cells will tend to clump when lifted from the dish. Pipetting the cells up and down will tend to disrupt these clumps. More effective dispersion of the clumps can be obtained by forcing the cells through a 100-μM nylon mesh.

COMMENTARY

Background Information

COS cells

COS cells are African green monkey kidney cells (CV-1) that have been transformed with an origin-defective SV40 virus, which has integrated into COS cell chromosomal DNA. Therefore, COS cells produce wild-type SV40 large T antigen but no viral particles. Since SV40 large T antigen is the only viral protein required in *trans* (i.e., its coding sequence need not be located on the DNA molecule on which it acts) for viral replication, SV40 origin–containing plasmids replicate in these cells to a high copy number (10,000 to 100,000 copies/cell) 48 hr posttransfection. If the plasmid carries a cDNA or genomic insert encoding a desired protein (under the control of the appropriate promoter), COS cells will express the protein at relatively high levels over a short period of time. Transfected COS cells produce protein in a burst that starts ~24 hr posttransfection and can last for up to a week. However, due to the excessive burden placed on the transfected cell by the replicating plasmid and the high levels of protein production, the transfected cells typically either die or lose the plasmid a week after transfection.

COS cells were developed by Yakov Gluzman (1981) as a host for the propagation of SV40 virus early-region mutants. The first SV40 origin–containing plasmids to be used in conjunction with COS cells were made by Lusky and Botchan (1981). Short-term expression systems using both COS cells and SV40 origin–containing plasmids were initially used to identify DNA sequences required for transcription of the human α1-globin gene (Mellon et al., 1981).

COS cells were first used to produce cell-surface and secreted proteins by Rose and Bergmann (1982), who looked at the expression of wild-type and mutant vesicular stomatitis virus glycoprotein in transfected cells. This technique was subsequently used to study the expression of insulin (Laub and Rutter, 1983), somatostatin (Warren and Shields, 1984), and acetylcholine receptors (Mishina et al., 1984). These experiments demonstrated that COS cells could be used to express biologically active cell-surface and secreted proteins. Furthermore, these proteins were correctly processed although they are normally not produced by COS cells.

COS cell expression was initially used to screen a cDNA library to isolate a cDNA encoding human granulocyte/macrophage colony-stimulating factor (Wong et al., 1985). This was subsequently extended to isolation of cDNAs encoding cell-surface proteins (Seed and Aruffo, 1987; Aruffo and Seed, 1987a) and DNA-binding proteins (Tsai et al., 1989).

The expressed protein produced in COS cells is, in most cases, biologically active. However, although COS cells are able to carry out some posttranslational modifications, they may not modify the expressed protein in exactly the same way as the cell that would normally produce it. For example, COS cells do not express the α-(1,3)fucosyltransferase which is capable of transferring fucose to either sialyl or asialyl precursors (Goelz et al., 1990). In addition, insufficient posttranslational modification occurs in the case of lymphocyte cell-surface

proteins, which tend to be underglycosylated in COS cells (Aruffo and Seed, 1987b). This might be due to an overburdening of the COS cell glycosylation machinery by the high levels of protein expression and/or by the lack of enzymes required to carry out the full posttranslational modifications.

An alternative to COS cells is provided by WOP cells (Dailey and Basilico, 1985), which are mouse 3T3 cells transformed by an origin-defective polyoma virus. Like COS cells, they produce no viral particles. However, they produce polyoma large T antigen and are therefore capable of replicating a plasmid containing a polyoma origin of replication to a copy number that is typically ten times lower than that obtained in COS cells. WOP cells are also more delicate than COS cells, making them harder to transfect. For these two reasons, COS cells should be used whenever possible. However, WOP cells should be used in those cases where a monoclonal antibody which is used to identify and/or purify the protein being expressed cross-reacts with COS cell proteins. This situation has arisen when transient expression in COS cells was carried out in order to clone a human protein with a mouse anti-human monoclonal antibody that also recognizes the equivalent monkey protein (Seed et al., 1987). In this situation, the mouse cell line presented a useful alternative to COS cells by avoiding monoclonal antibody cross-reactivity.

Vectors

The main requirements of any COS cell expression/shuttle vector are: (1) an SV40-derived origin of replication, (2) appropriate eukaryotic transcription regulatory elements (i.e., enhancer, promoter, and polyadenylation signal sequences), (3) a prokaryotic origin of replication, and (4) a prokaryotic genetic marker for selection in *Escherichia coli*. A particularly useful example of such a vector is CDM8 (Fig. 16.13.1; Seed, 1987). The eukaryotic transcription element of CDM8 is composed of the cytomegalovirus (CMV) enhancer-promoter, with an SV40 virus–derived intron and polyadenylation signal; the CMV promoter is a *cis* element (i.e., one that is located adjacent to the DNA it acts on) that directs transcription of the DNA subcloned downstream from it. The prokaryotic genetic

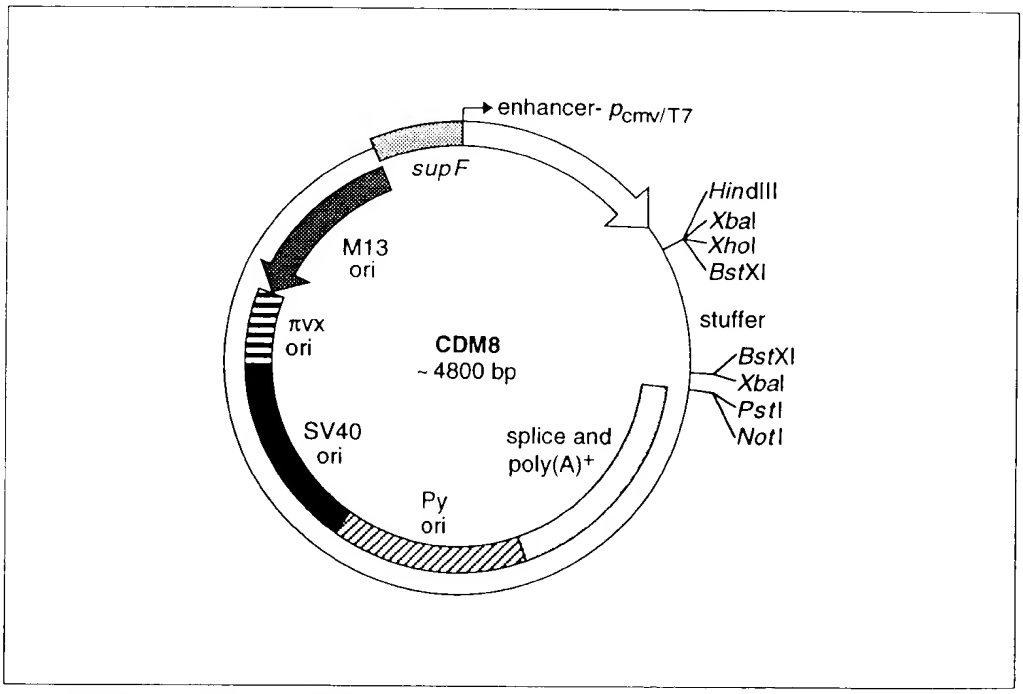


Figure 16.13.1 CDM8. CDM8 (Seed, 1987) is a COS cell expression/shuttle vector that contains an SV40-derived origin of replication (SV40 ori), eukaryotic transcription regulatory elements [splice and poly(A)$^+$], a prokaryotic origin of replication (πvx; derived from the pBR322 ori), and a prokaryotic genetic marker (*supF*). In addition, CDM8 contains an M13 origin of replication (M13 ori), a T7 RNA polymerase promoter ($p_{CMV/T7}$), and a polyoma-derived origin of replication (Py ori). Any restriction endonuclease sites shown can be used for cloning, but the inserted fragment must have its 5′ end nearest the enchancer $p_{CMV/T7}$. The stuffer sequence is used to detect a size difference after a restriction digest.

marker in CDM8 is provided by the *supF* (amber suppressor) gene. CDM8 is propagated in host bacterial cells containing helper plasmid P3, which contains amber mutations in the genes responsible for tetracycline and ampicillin resistance (*UNIT 1.4*). (P3 has been introduced into many *E. coli* strains.) When CDM8 is transformed into an *E. coli* strain containing P3 (*UNIT 1.8*), the amber mutations are suppressed, rendering the host resistant to tetracycline and ampicillin. In addition to these elements, the CDM8 expression/shuttle vector contains an M13 origin of replication so that it can be used for the production of single-stranded DNA (*UNIT 1.15*), a T7 RNA polymerase promoter for preparation of mRNA in vitro (*UNIT 3.8*), and a polyoma virus–derived origin of replication which permits plasmid replication in WOP cells.

Other vectors commonly used for COS cell transient expression include pXM (Yang et al., 1986) and pDC201 (Sims et al., 1988). These two plasmids contain the adenovirus-2 major late promoter and tripartite mRNA leader. This element acts in conjunction with the adenovirus VA RNA (also produced by the vectors) to increase the translatability of the mRNA encoding the desired protein. It is thought that adenovirus VA RNAs increase translation efficiency of mRNAs containing the major late promoter tripartite leader by facilitating the interaction between mRNA and a 43S ribosomal protein translation preinitation complex (Kaufman, 1985).

Critical Parameters

Efficiency of transfection depends critically on the length of time that the cells (COS or other cells) are incubated in the presence of DEAE-dextran/DNA. Longer periods of time result in higher transfection efficiencies. However, the DEAE-dextran/chloroquine solution is quite toxic to cells and in general, cells should not be in its presence for >4 hr. In the past, DEAE-dextran transfections were carried out in the absence of serum, because a precipitate of unknown composition that seemed to be very toxic formed in DEAE-dextran/calf serum mixtures. Medium containing 10% NuSerum, on the other hand, does not form this precipitate and tends to enhance the ability of the cells to tolerate DEAE-dextran; thus NuSerum should always be included in transfection medium.

Efficiency of transfection can also be affected by the quality of the DNA and the age of the DEAE-dextran/chloroquine solution. It is preferable to use CsCl-purified or other highly purified DNA whenever possible (*UNIT 1.7*). However, miniprep DNA (*UNIT 1.6*) or DNA purified using a pZ523 column (5′→3′) or by other methods can also be used. The DEAE-dextran chloroquine solution can be kept at 4°C for several months but it is wise to prepare it fresh about every 3 months.

COS cells can be obtained from the American Type Culture Collection; several sublines exist including COS-1 and COS-7. The COS-7 subline is recommended because it produces a higher plasmid copy number. These cells grow as a monolayer in DMEM-10 CS in a humidified 37°C, 5% CO_2 incubator; however, the ATCC grows its COS cells in DMEM-10 fetal calf serum (FCS). Since FCS is significantly more expensive than CS, changing the growth medium is worthwhile. The change from FBS to CS should be done slowly over 1 to 2 weeks.

Because the growth characteristics, transfectability, and protein expression properties of COS cells change with time and with repeated subculturing, and because these changes tend not to favor the production of high levels of proteins, it is prudent to freeze aliquots of the original COS cell stock in DMEM-10 CS/10% DMSO for later use in a −70°C freezer for 24 hr and then transfer them to a −150°C (liquid nitrogen) freezer. COS cells grow rapidly requiring passage every 4 to 5 days; typically, a confluent plate of cells is split 1 to 10.

To obtain good levels of transient protein production from transfected COS cells, it is very important to replate the transfected cells onto new dishes with fresh medium the morning after tranfection. In addition to enhancing protein production, replating the transfected cells allows lifting of the cells from the dish using only PBS/EDTA (without trypsin). This is very important when transient expression by COS cells is used to produce cell-surface proteins.

Anticipated Results

The basic protocol should yield transfection efficiencies of 40% to 70%. When COS cells are being used to produce cell-surface or intracellular proteins, it can be expected that each transfected cell will express several thousand copies of this protein (10,000 to 100,000 copies/cell) 72 hr posttransfection. If COS cells are used to produce secreted proteins, up to 1 µg/ml of protein can be recovered from the supernatant of a 100-mm dish of transfected cells 1 week posttransfection. However, the amount of protein produced by COS cells can vary dramatically depending on the protein being pro-

duced. This was the case when COS cells were used to produce soluble immunoglobulin fusion forms of cell-surface proteins (Aruffo et al., 1990). In this case, one of the fusion proteins, CD8 immunoglobulin, was secreted from COS cells at high levels (1 μg/ml) while the other, CD44 immunoglobulin, was secreted very poorly if at all. It was found that the CD44 fusion protein was sequestered inside the cell. To obtain efficient secretion of the CD44 fusion protein, it was necessary to change the amino-terminal signal sequence of the CD44 fusion protein. Interestingly, the native cell-surface forms of both CD8 and CD44 are expressed equally efficiently on the surface of transfected COS cells. For some of these immunoglobulin fusion proteins, 0.5 ml of medium contained plenty of protein (~500 ng) after concentration using a protein A–affinity matrix. In some cases, it is possible to use such COS cell supernatants directly without further purification (Aruffo et al., 1990).

Time Considerations

It is important not to transfect the cells too soon after replating; >8 to 12 hr should pass between the time the cells are seeded on the plate and the time of transfection. Once the transfection has started, the DEAE-dextran/DNA mixture should be left on the cells for a minimum of 2 hr and a maximum of 4 hr; because the mixture increases transfection efficiency it should remain in contact with the cells as long as they appear viable. After transfection, the cells look quite unhealthy and 12 to 24 hr posttransfection, they should be replated in new dishes with fresh medium.

The peak of plasmid replication in transfected COS cells occurs 48 to 72 hr posttransfection. Protein production starts 24 hr posttransfection but peaks 72 to 96 hr posttransfection. Thus, when expressing cell-surface or cytoplasmic proteins, the cells should be harvested 72 to 96 hr posttransfection. However, transfected cells continue to produce protein for up to a week posttransfection and when expressing secreted proteins, the supernatants should be harvested a week posttransfection.

Literature Cited

Aruffo, A. and Seed, B. 1987a. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl. Acad. Sci. U.S.A.* 84:8573-8577.

Aruffo, A. and Seed, B. 1987b. Molecular cloning of two CD7 (T cell leukemia antigen) cDNAs by a COS cell expression system. *EMBO J.* 6:3313-3316.

Aruffo, A., Stamenkovic, I., Melnick, M., Underhill, C.B., and Seed, B. 1990. CD44 is the principal cell-surface receptor for hyaluronate. *Cell* 61:1303-1313.

Dailey, L. and Basilico, C.J. 1985. Sequences in the polyoma virus DNA regulatory region involved in viral DNA replication and early gene expression. *J. Virol.* 54:739-749.

Gluzman, Y. 1981. SV-40 transformed simian cells support the replication of early SV40 mutants. *Cell* 23:175-182.

Goelz, S.E., Hession, C., Goff, D., Griffiths, B., Tizard, R., Newman, B., Chi-Rosso, G., and Lobb, R. 1990. ELFT: A gene that directs the expression of an ELAM-1 ligand. *Cell* 63 1349-1356.

Kaufman, R.J. 1985. Identification of the components necessary for adenovirus translational control and their utilization in cDNA expression vectors. *Proc. Natl. Acad. Sci. U.S.A.* 82:689-693.

Laub, O. and Rutter, W.J. 1983. Expression of the human insulin gene and cDNA in a heterologous mammalian system. *J. Biol. Chem.* 258 6043-6050.

Lusky, M. and Botchan, M. 1981. Inhibition of SV40 replication in simina cells by specific pBR322 DNA sequences *Nature (Lond.)* 293:79-81.

Mellon, P., Parker, V., Gluzman, Y., and Maniatis, T. 1981. Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. *Cell* 27:279-288.

Mishina, M., Kurosaki, T., Tobimatsu, T., Morimoto, Y., Noda, M., Yamamoto, T., Terao, M., Lindstrom, J., Takahashi, T., Kuno, M., and Numa, S. 1984. Expression of functional acetylcholine receptor from cloned cDNAs. *Nature (Lond.)* 307:604-608.

Rose J.K. and Bergmann, J.E. 1982. Expression from cloned cDNA of cell-surface, secreted forms of the glycoprotein of vesicular stomatitis virus in eucaryotic cells. *Cell* 30.753-762.

Seed, B. 1987. An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor CD2. *Nature (Lond.)* 329:840-842.

Seed, B. and Aruffo, A. 1987. Molecular cloning of the CD2 antigen, the T cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. U.S.A.* 84:3365-3369.

Sims, J.E., March, C.J., Cosman, D., Widmer, M.B., MacDonald, H.R., McMahan, C.J., Grubin, C.E., Wignall, J.M., Jackson, J.L., Call, S.M., Friend, D., Alpert, A.R., Gillis, S., Urdal, D.L., and Dower, S.K. 1988. cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily. *Science* 241:585-589.

Tsai, F., Martin, D.I.K., Zon, L.I., D'Andrea, D., Wong, G.G., and Orkin, S.H. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mam-

malian cells. *Nature (Lond.)* 339:446-451.

Warren, T.G. and Shields, D. 1984. Expression of preprosomatostatin in heterologous cells: Biosynthesis, posttranslational processing, and secretion of mature somatostatin. *Cell* 39:547-555.

Wong, G.G., Witek, J.S., Temple, P.A., Wilkens, K.M., Leary, A.C., Luxenber, D.P., Jones, S.S., Brown, E.L., Kay, R.M., Orr, E.C., Shoemaker, C., Golde, D.W., Kaufman, R.J., Hewick, R.M., Wang, E.A., and Clark, S.C. 1985. Human GM-CSF: Molecular cloning of the complementary DNA and purification of the natural and recombinant proteins. *Science* 228:810-815.

Yang, Y., Ciarletta, A.B., Temple, P.A., Chung, M.P., Kovacic, S., Witek-Giannotti, J.S., Leary, A.C., Kriz, R., Donahue, R.E., Wong, G.G., and Clark, S.C. 1986. Human IL-3 (multi-CSF): Identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell* 47:3-10.

Yokoyama, W.M. 1991. Flow cytometry analysis using the Becton Dickinson FACScan. *In* Current Protocols in Immunology (J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, and W. Strober, eds.) pp. 5.4.1-5.4.15. Greene Publishing and Wiley-Interscience, New York.

Key Reference

Warren and Shields. 1984. See above.

This article shows that COS cells can be used as an efficient, short-term, mammalian epression system for the production of proteins.

Contributed by Alejandro Aruffo
Bristol-Myers Squibb
Seattle, Washington

CURRENT PROTOCOLS IN MOLECULAR BIOLOGY

VOLUME 2

John Wiley & Sons, Inc.

CORE 13 (S36)

Library of Congress Cataloging in Publication Data:

Current protocols in molecular biology. 3 vols.

1. Molecular biology—Technique. 2. Molecular biology—Laboratory manuals. I. Ausubel, Frederick M.

QH506.C87 1987 574.8'8'028 87-21033

ISBN 0-471-50338-X

Printed in the United States of America

20 19 18 17 16 15 14 13

pTrc 99 A Expression Vector Kit

- *For expression of proteins encoded by inserts lacking a start codon.*

pTrc 99 A is a derivative (1) of pKK233-2 expression vector with a strong *trc* promoter upstream of the MCS and a strong *rrn*B transcription termination signal downstream. It contains an *Nco* I site next to the *trc* promoter for ligation and direct expression of inserts possessing the start codon ATG. *Nco* I linkers provided with the kit permit ligation of inserts lacking a start codon and subsequent expression in all three possible reading frames. The kit includes the following components:

pTrc 99 A Expression Vector
Nco I Linker 5'-d[CAGCCATGGCTG]-3'
Nco I Linker 5'-d[AGCCATGGCT]-3'
Nco I Linker 5'-d[GCCATGGC]-3'
E. coli JM105

Induction: *trc* promoter is inducible with 1-5 mM IPTG. It contains the *trp* -35 region and *lac* UV5 -10 region separated by 17 bp. *trc* promoter is very strong; even uninduced cells may show a low level of expression.

Expression: *Nco* I linker supplies ATG start codon in all three reading frames for expression of insert. Vector provides ribosome-binding site.

Host(s): Supplied with *E. coli* JM105. Plasmid provides *lac* I^q repressor.

Selectable marker(s): Plasmid confers resistance to 100 µg/ml ampicillin.

Amplification: Recommended.

Control Regions: **Expression control region:** *trc* promoter: -10: 216-222; -35: 193-198; Ribosome binding site: 255. **MCS:** 265-314. ***rrn*B operon region:** 5S rRNA region: 409-528; *rrn*B T1 terminator: 532-575; *rrn*B T2 terminator: 707-734. **β-lactamase gene region:** Promoter: -10: 799-804; -35: 776-781; Start codon (ATG): 846; Stop codon (TAA): 1704. ***lacI*^{q} gene region:** Start codon (GTG): 3055; Stop codon (TGA): 4135. **Plasmid replication region:** Site of replication initiation: 2465-2467; Region necessary for replication: 1771-2467.

ORDERING INFORMATION

Product	Quantity	Code No	Price (US $)
pTrc 99 A Expression Vector Kit	1 kit	27-5007-01	211.00

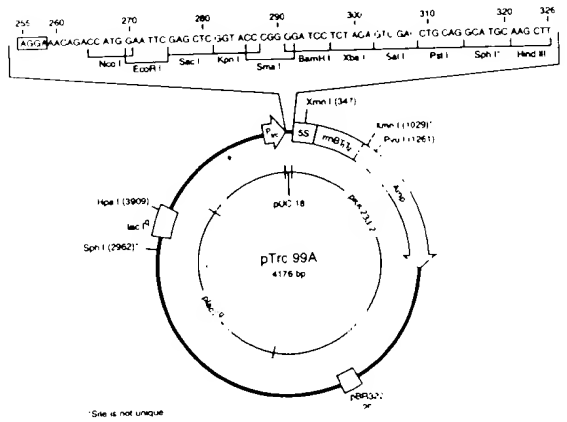
Companion Products	Code No.	Page
Nco I	27-0971-01	131

Reference

1. Amann, E. *et al.*, *Gene* 69, 301 (1988).

Note: *For research purposes only. Commercial use requires a license from Behringwerke Aktiengesellschaft, 3550 Marburg, Germany.*

■ / GenBank Accession Number U13872.



pKK223-3 Expression Vector

- *For over-expression of proteins under the control of the strong* tac *promoter.*

pKK223-3 (1) contains the strong *tac* promoter upstream from the multiple cloning site and the strong *rrn*B ribosomal terminator downstream for control of protein expression (2-4).

Induction: *tac* promoter is inducible with 1-5 mM IPTG. Promoter is very strong; even uninduced cells may show a low level of expression.

Expression: Genes containing a ribosome-binding site and ATG can be inserted into any unique site in the MCS for expression. The ribosome-binding site on the plasmid can be utilized to express inserts if the ATG start codon of the insert is within 5-13 bp from the provided ribosome-binding site (5, 6).

Transcription terminators: *rrn*B transcription terminators stabilize the plasmid by inhibiting read-through transcription initiated from the *tac* promoter in the parent plasmid.

Host(s): *lac* I^q strains; *E. coli* JM105 is recommended.

Selectable marker(s): Plasmid confers resistance to 100 µg/ml ampicillin.

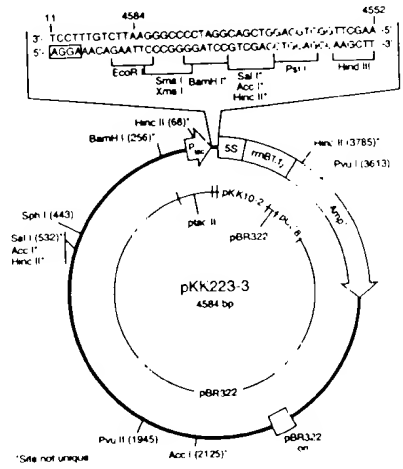
Amplification: Recommended.

Control Regions: **Expression control region:** *tac* promoter: -10: 50-44; -35: 72-67; Ribosome binding site: 11. **MCS:** 4552-4587. ***rrn*B operon region:** 5S rRNA region: 4471-4352; *rrn*B T1 terminator: 4348-4305; *rrn*B T2 terminator: 4173-4146. **β-lactamase gene region:** Promoter: -10: 4082-4077; -35: 4105-4100; Start codon (ATG): 4035; Stop codon (TAA): 3178. **Plasmid replication region:** Site of replication initiation: 2417-2415; Region necessary for replication: 3111-2415.

ORDERING INFORMATION

Product	Quantity	Code No.	Price (US $)
pKK223-3 Expression Vector	25 µg	27-4935-01	135.00

■ / GenBank Accession Number M77749.



References

1. Brosius, J. and Holy, A., *Proc. Natl. Acad. Sci. USA* 81, 6929 (1984).
2. Amann, E. *et al.*, *Gene* 25, 167 (1983).
3. Frost, J. W. *et al.*, *Biochemistry* 23, 4470 (1984).
4. Jaffe, B. M. *et al.*, *Biochemistry* 27, 1869 (1988).
5. Kozak, M., *Microbiol. Rev.* 47, 1 (1983).
6. Gold, L., *Ann. Rev. Biochem.* 57, 199 (1988).

pSVK 3 Expression Vector

- *For* in vivo *expression in mammalian cell lines (1).*

The phagemid pSVK 3 is a true shuttle vector, containing both the *colE* 1 origin for replication in *E. coli* and the origin of replication, early promoter, mRNA splice site and polyadenylation signals from SV40 for replication and expression in eukaryotic cells. The presence of the f1 origin of replication permits single-stranded phage production following infection of bacterial cultures with the helper phage M13KO7.

Expression: Expression is controlled by the uninducible SV40 early promoter. Expression is transient, but can be stable if the gene insert is selectable.

Sequencing: Both double-stranded and single-stranded [(+) strand] sequencing are possible. A protocol for production of single-stranded DNA is provided with the vector.

***In vitro* transcription:** A T7 RNA polymerase promoter allows for *in vitro* transcription of cloned inserts.

Host(s): *E. coli* and mammalian cells (CV-1, COS and HeLa cells).

Selectable marker(s): Plasmid confers resistance to 100 µg/ml ampicillin in *E. coli.*

Control Regions: SV40 origin and early promoter region: Minimal region for replication: 239-314; TATA box: 252-258; 21-bp repeats 170-232, 72-bp repeats: 23-166; Transcription start region: 279-286. MCS: 377-435. SV40 early splice region: Small T antigen intron: 630-695. **SV40 polyadenylation region:** PolyA signal sequence: 1331-1336; PolyA site: 1352. **T7 promoter:** 352-374. **β-lactamase gene region:** Promoter region: -10: 2017-2022; -35: 1994-1999; Start codon (ATG): 2064; Stop codon (TAA): 2922. **Plasmid replication region:** Site of replication initiation: 3683-3685; Region necessary for replication: 2989-3685. **f1 region:** 1478-1932; Origin of f1 replication: 1770.

Reference

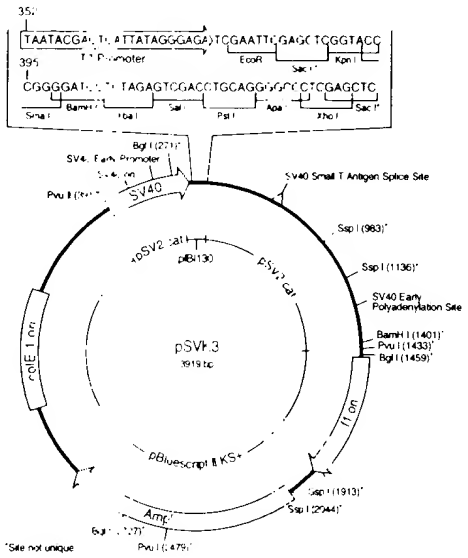
1. Mongkolsuk, S., *Gene* 70, 313 (1988).

ORDERING INFORMATION

Product	Quantity	Code No.	Price (US $)
pSVK 3 Expression Vector	25 µg	27-4511-01	154.00

Companion Products	Code No.	Page
M13KO7 Helper Phage	27-1524-01	98

■ / GenBank Accession Number U13867.



pSVL SV40 Late Promoter Expression Vector

- *For high-level transient expression in eukaryotic cells.*

Expression: Genes inserted into the multiple cloning site are transiently expressed from the first ATG using the SV40 late promoter (1, 2). Transcripts are spliced and polyadenylated using the SV40 VP1 processing signals. Absence of the pBR322 "poison sequence" results in higher plasmid copy number and higher level of transient expression in COS cells (3).

Host(s): Any suitable restriction-minus *E. coli* strain for amplification and cloning; mammalian cells including COS (3), CV1, HeLa and 3T3 cells. Highest level of eukaryotic expression is found in T antigen-producing cells, such as COS.

Selectable marker(s): Plasmid confers resistance to 100 µg/ml ampicillin in *E. coli.* Stable transformants in eukaryotes can be isolated if the inserted gene is selectable.

Amplification: Recommended.

Control Regions: SV40 origin and late promoter region: Minimal region for replication: 4824-4834; 21-bp repeats: 41-103; 72-bp repeats: 107-250; Major transcription start site: 325. **VP1 region:** VP1 leader sequence: 323-525; VP1 intron: 526-1493 MCS: 1506-1535. **SV40 polyadenylation region:** PolyA signal sequence: 1654-1659; PolyA site: 1671. **β-lactamase gene region:** Promoter region: -35: 4230-4225; -10: 4207-4202; Start codon (ATG) 4160; Stop codon (TAA): 3302. **Plasmid replication region:** Site of replication initiation: 2541-2539; Region necessary for replication: 3235-2539.

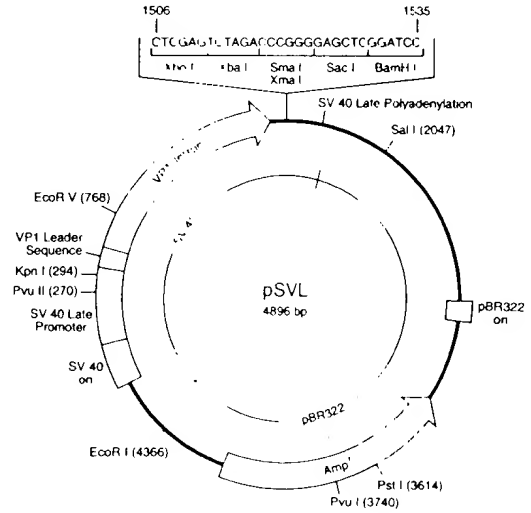
References

1. Sprague, J. *et al., J. Virology* 45, 773 (1983).
2. Templeton, D. and Eckhart, W., *Mol. Cell. Biol.* 4, 817 (1984).
3. Lusky, M. and Botchan, M., *Nature* 293, 79 (1981).

ORDERING INFORMATION

Product	Quantity	Code No.	Price (US $)
pSVL SV40 Late Promoter Expression Vector	25 µg	27-4509-01	135.00

■ / GenBank Accession Number U13868.



pMSG Inducible Mammalian Expression Vector

- *For inducible expression in mammalian cells (1).*

Induction: Expression is controlled by the MMTV LTR promoter (mouse mammary tumor virus long terminal repeat) located upstream of the multiple cloning site (MCS). Genes inserted into the MCS are translated from the first ATG of the insert and expressed from the promoter, which is glucocorticoid-inducible (0.1 µM dexamethasone).

Selection in eukaryotes: pMSG contains the *E. coli* xanthine-guanine phosphoribosyltransferase (*gpt*) gene which is expressed from the SV40 early promoter and therefore provides a marker for selecting stable transformants in eukaryotic cells (1-3). Cells with a functional *gpt* gene are resistant to mycophenolic acid.

Stable expression: Cells must be hormonally responsive (1-6) for expression from the MMTV promoter. pMSG will integrate randomly into the host genome under selective pressure of mycophenolic acid.

Transient expression: After transfection, transient expression is inducible with dexamethasone. Cells must be hormonally responsive (1-6) for expression from the MMTV promoter.

Host(s): *recA*⁻ *E. coli* strains such as HB101 (Note: pMSG contains regions of internal homology; use with some bacterial strains may lead to deletions in the vector.) Mammalian cells including Chinese hamster ovary (CHO) (1), HeLa cells (3) and mouse Ltk⁻ (4) and 3T6 (5). For induction, cells must be responsive to glucocorticoid compounds.

Selectable marker(s): Plasmid confers resistance to 100 µg/ml ampicillin in *E. coli*. Stably transformed mammalian cells are resistant to mycophenolic acid.

Amplification: Recommended.

Control Regions: MCS: 1-23. MMTV LTR region: LTR: 6178-7492; Hormone response element: 7122-7306; Approximate site for transcription initiation: 7358. SV40 early splice region: Small T antigen intron: 101-165. SV40 polyadenylation region: PolyA signal sequence: 801-806; PolyA site: 822. SV40 origin and early promoter region: Minimal region for replication: 1116-1191; TATA box: 1129-1135; 21-bp repeats: 1047-1109; 72-bp repeats: 900-1043; Transcription initiation region: 1156-1163. *gpt* gene region: *gpt* gene: 1218-2279; Start codon (ATG): 1416, Stop codon (TAA): 1872. SV40 early splice region: Small T antigen intron: 2351-2415. SV40 polyadenylation region: PolyA signal sequence: 3051-3056; PolyA site: 3072. β-lactamase gene region: Promoter region: -10: 4036-4041; -35: 4013-4018; Start codon (ATG): 4083; Stop codon (TAA): 4941. Plasmid replication region: Site of replication initiation: 5702-5704; Region necessary for replication: 5008-5704.

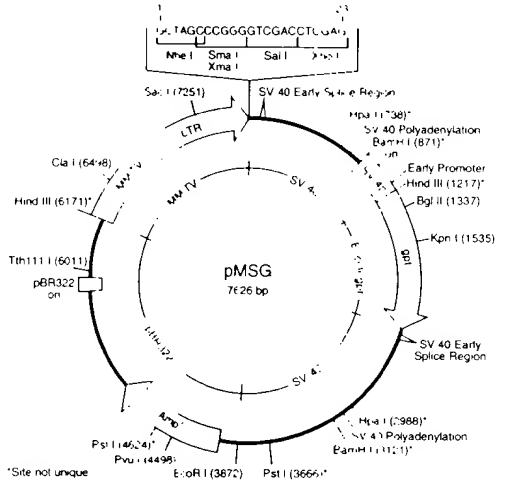
ORDERING INFORMATION

Product	Quantity	Code No.	Price (US $)
pMSG Inducible Mammalian Expression Vector	25 µg	27-4506-01	135.00

References

1. Lee, F. *et al., Nature* 294, 228 (1981).
2. Mulligan, R. and Berg, P., *Proc. Natl. Acad. Sci. USA* 78, 2072 (1981).
3. Klessig, D. F. *et al., Mol. Cell. Biol.* 4, 1354 (1984).
4. Hall, C. V. *et al., J. Mol. Appl. Gen.* 2, 101 (1983).
5. Chapman, A. B. *et al., Mol. Cell. Biol.* 3, 1421 (1983).
6. Morley, K. L. *et al., Nucl. Acids Res.* 15, 6973 (1987).

■ / GenBank Accession Number U13860.



pCH110 Eukaryotic Assay Vector

- *For monitoring and normalizing expression in eukaryotic cells (1).*

Expression: pCH110 contains a functional *lacZ* gene which is expressed from the SV40 early promoter in eukaryotes and from the *E. coli gpt* promoter in prokaryotes (2). When co-transfected with a second plasmid, pCH110 provides an internal marker to monitor and normalize transient and stable expression between experiments in eukaryotic cells (3).

Promoter screening: Insertion of a promoter into the *Hind* III site allows monitoring of β-galactosidase expression (1, 4).

Host(s): *E. coli* and mammalian cells (including COS monkey cells, and mouse Ltk⁻ cells)

Selectable marker(s): Plasmid confers resistance to 100 µg/ml ampicillin in *E. coli*.

Amplification: Recommended.

Control Regions: SV40 origin and early promoter region: Minimal region for replication: 7098-7023; TATA box: 7036-7042, 21-bp repeats: 6954-7017; 72-bp repeats: 6807-6950; Transcription start region: 7063-7070. *gpt* gene region: *gpt* gene: 7-208; TATAA: 13-17, Start codon (ATG): 85. *trpS* gene region: 209-292. *lacZ* gene region: *lacZ* gene: 295-3605; Stop codon (TAA): 3340. SV40 polyadenylation region: PolyA signal sequence: 3666-3671; PolyA site: 3687. β-lactamase gene region: Promoter region: -10: 4651-4656; -35: 4628-4633; Start codon (ATG): 4698; Stop codon (TAA): 5556. Plasmid replication region: Site of replication initiation: 6317-6319; Region necessary for replication: 5623-6319.

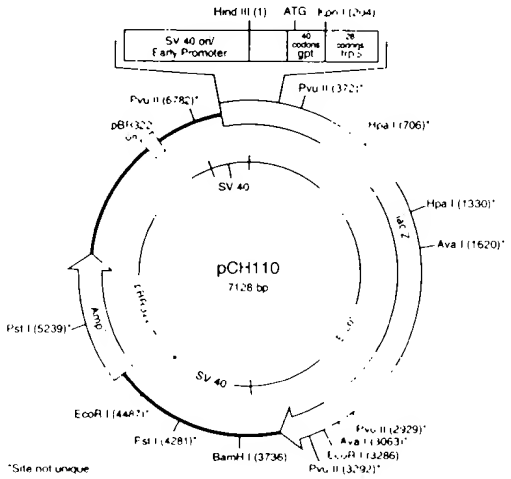
References

1. Miller, J. H. in *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory) (1972), p. 352.
2. Hall, C. V. *et al., J. Mol. Appl. Gen.* 2, 101 (1983).
3. Herbomel, P. *et al., Cell* 39, 653 (1984).
4. Lee, F. *et al., Nucl. Acids Res.* 12, 4191 (1984).

ORDERING INFORMATION

Product	Quantity	Code No	Price (US $)
pCH110 Eukaryotic Assay Vector	25 µg	27-4508-01	135.00

■ / GenBank Accession Number U13845.



DIRECT
A Trademark of Pharmacia Biotech

biodirectory™

Pharmacia Biotech

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 267, No. 9, Issue of March 25, pp. 6291-6296, 1992
Printed in U.S.A.

Epitope Mapping of Anti-recA Protein IgGs by Region Specified Polymerase Chain Reaction Mutagenesis*

(Received for publication, October 14, 1991)

Masayuki Ikeda, Kaoru Hamano, and Takehiko Shibata‡

From the Laboratory of Microbiology, RIKEN (The Institute of Physical and Chemical Research), Wako-shi, Saitama 351-01, Japan

Monoclonal IgGs were shown to be useful for the specific inhibition of a set of activities of the recA protein, a key protein in homologous genetic recombination. The mapping of the epitopes for these IgGs and site-directed mutagenesis based on the mapping will facilitate location of the functionally active sites on the tertiary structure of the protein, which is being solved by means of physicochemical techniques. We developed a novel technique for region-specified mutagenesis and applied the technique to epitope mapping. Using the polymerase chain reaction in the presence of deoxyinosine triphosphate, we introduced random base substitutions specifically into a region of the *recA* gene defined by a pair of primers. RecA mutants exhibiting altered antigenicity were selected, in plaque-immunoblotting experiments, from libraries of mutagenized *recA* genes constructed on the λgt11 expression vector. Mutant recA genes were obtained at the frequency of about 10^{-2} among the plaques expressing fused recA genes and then each one was expressed as a whole protein, which was characterized by enzyme-linked immunosorbent assay. Analyzing the DNA sequences of the mutant *recA* genes, we located at the amino acid sequence level the epitopes for two anti-recA IgGs which could not be located in previous studies. One of the antibodies was shown to prevent self-assembly of the recA protein and the other was suggested to inhibit the binding of double-stranded DNA. Thus, the active sites involved in these functions would be located in the space around or near the relevant epitope.

The recA protein, and its prokaryotic and virus (T4 phage) analogues promote "homologous pairing" and "strand exchange" between homologous double-stranded and single-stranded DNAs through ATP (or dATP)-dependent reactions *in vitro*, and were shown to play an essential role in homologous genetic recombination *in vivo*. Homologous pairing is the formation of an intermolecular duplex ("heteroduplex") between a couple of homologous single-stranded and double-stranded DNAs, and strand exchange is the processing of the heteroduplex, such as its elongation. Each of these reactions consist of a number of substeps and the recA protein or its analogues appear to have various active sites that promote each of these substeps, such as an ATP-binding site, ATPase catalytic center, binding site for single-stranded DNA, binding site for double-stranded DNA, and sites for self-polymerization. The localization of these active sites on the tertiary structure of the recA protein is essential for understanding the mechanisms of the underlying biochemical functions of the protein. However, only sites related to ATPase have been partly located at the amino acid sequence level. The mapping of mutation sites as well as x-ray crystallographic analysis of the protein are the main means to this end. A series of our studies involving the use of anti-recA protein monoclonal IgGs is also an approach to the same goal (see Shibata *et al.*, 1991, for review).

We have isolated clones of mouse hybridomas which produce anti-recA protein IgGs (Makino *et al.*, 1985). Two (ARM193 and ARM191) of these anti-recA protein IgGs each inhibit a set of activities of the recA protein without affecting the others; *i.e.* ARM193 severely inhibits the unwinding of the double helix and strand exchange, but allows homologous pairing and single-stranded DNA-dependent ATP hydrolysis (Ikawa *et al.*, 1989; Makino *et al.*, 1985, 1987). On the other hand, ARM191 inhibits the homologous pairing and unwinding of the double helix, but only affects the single-stranded DNA-dependent ATP hydrolysis a little (Makino *et al.*, 1985). ARM193 was suggested to affect the site for the interaction between recA polypeptides (Ikawa *et al.*, 1989) and ARM191 to affect the site on the recA polypeptide for the binding to double-stranded DNA (Makino *et al.*, 1985). We preliminarily located the epitopes for both ARM193 and ARM191 in a C-terminal 88 amino acid region (Phe^{260}-Glu^{347}) of the recA polypeptide by examining the cross-reaction of proteolytic fragments. However, we failed to map them more precisely, since none of the subfragments of the 88-amino acid region exhibited significant cross-reaction with either of the IgGs (Ikeda *et al.*, 1990). Therefore, it was necessary to introduce another technique to overcome the problem. Here, we describe a novel technique for region-specified mutagenesis and, as an application of this technique, the mapping of the epitopes of ARM193 and ARM191 at the amino acid sequence level in distinct but slightly overlapped subregions in the C-terminal 88-amino acid region.

MATERIALS AND METHODS

recA Protein - The purified recA protein was fraction V prepared as described (Shibata *et al.* 1981).

Oligonucleotides - Oligonucleotides were synthesized with a DNA synthesizer (Du Pont-New England Nuclear CODER300) and purified with NENSORB PREP (Du Pont-New England Nuclear).

Techniques for Cloning of DNAs - Treatment with restriction endonucleases and DNA ligase, and the isolation and cloning of the DNA fragments onto vectors were carried out as described (Berger and Kimmel, 1987; Maniatis *et al.*, 1982).

Antibodies and Immunochemical Techniques - The anti-recA pro-

* This research was supported by a grant from the Biodesign Research Program from RIKEN (The Institute of Physical and Chemical Research). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Laboratory of Microbiology, RIKEN, Wako-shi, Saitama 351-01, Japan. Fax: 81-484-62-4671.

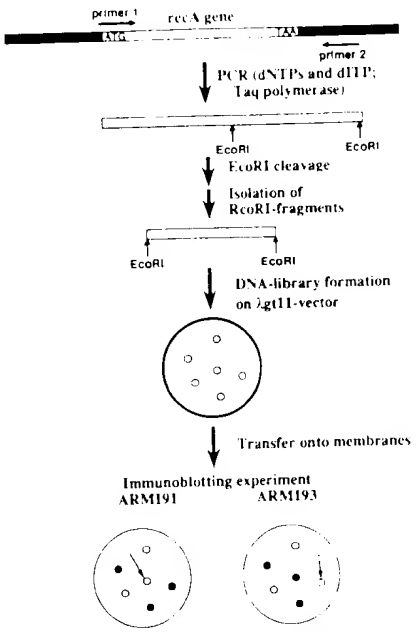


FIG. 1. **Region-specified PCR mutagenesis.** DNA encoding the *E. coli recA* gene flanked by primers 1 and 2 was amplified by PCR in the presence of dITP with the use of *Taq* DNA polymerase. *Eco*RI fragments of the amplified DNA which encoded the C-terminal region of the recA protein were cloned on a λgt11 expression vector to construct DNA libraries of the mutagenized *recA* genes. With appropriate orientation of a fragment relative to the vector, the C-terminal region of the *recA* gene was connected to the *lacZ* gene in-frame. Proteins expressed in the plaques obtained from the libraries were transferred to a pair of membranes and then the cross-reaction with either anti-recA protein IgG ARM191 or ARM193 was tested. The plaques showing cross-reaction with only one of the IgGs (indicated by *arrows*) were picked up and subjected to further cross-reaction tests. The *closed circles* in the *big circles* at the bottom of the figure denote plaques which showed cross-reaction with the indicated IgG.

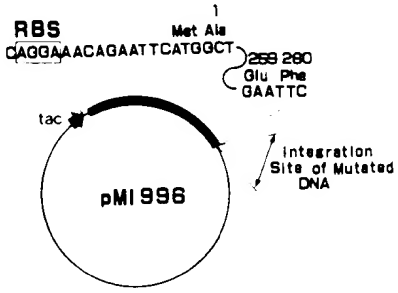


FIG. 2. **Structure of pMI996 for expression of mutated *recA* genes.** DNA encoding the N-terminal region of the recA polypeptide was put under the control of the *Tac* promoter on a multicopy plasmid (pKK223-3). *Eco*RI fragments of the mutagenized *recA* gene were inserted at the *Eco*RI site at the codons for Glu^{259}-Phe^{260}.

tein monoclonal IgGs, ARM191, ARM193, and ARM414, were described previously. We used affinity purified preparations of these IgGs. An anti-recA protein monoclonal IgG, MAb156, was isolated by Karu and Allen (Karu and Allen, 1982), and a purified preparation of this IgG was provided by Dr. Alexander Karu (University of California, Berkeley) and Dr. A. John Clark (University of California, Berkeley).

The enzyme-linked immunosorbent assay (ELISA)[1] was carried out as described previously. Unless otherwise stated, samples of cell-free extracts were diluted in PBS (50 mM potassium phosphate buffer (pH 7.2) containing 150 mM NaCl) and put into the sample wells o a microtiter plate with 96 wells. On the other hand, the concentratio of the anti-recA protein IgG was first adjusted to 30 μg of protein/m ("no dilution") in PBS Tween (PBS containing 0.05% Tween 20) an then a series of dilutions was prepared in PBS/Tween.

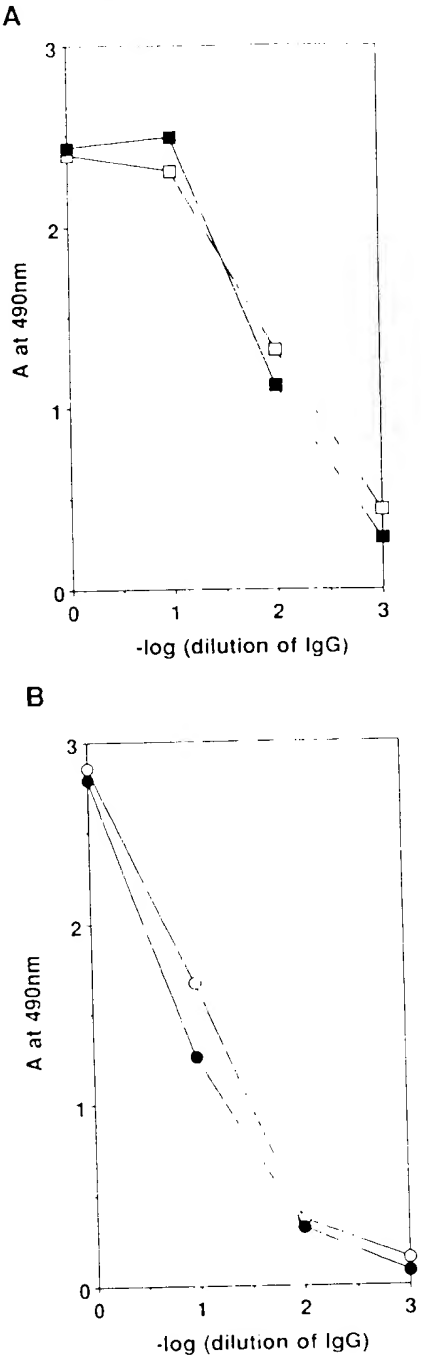


FIG. 3. **Absence of competition in cross-reaction between anti-recA protein IgGs, ARM191 and ARM193.** The wells of a microtiter plate were coated with the purified recA protein. Then, a solution of a tested anti-recA protein (nondilution: 1.5 μg/ml) containing the other IgG was put into the wells and allowed to cross-react with the recA protein. ARM191 belongs to IgG_1 and ARM193 to IgG_{2b}. Thus, the amounts of ARM191 and ARM193 bound to the recA protein were specifically measured by ELISA with the use of an appropriate subclass-specific antibody. *A*: □, the binding of ARM191 in the absence of ARM193; ■, the binding of ARM191 in the presence of 5 μg of ARM193/ml. *B*: ○, the binding of ARM193 in the absence of ARM191; ●, the binding of ARM193 in the presence of 5 μg o ARM191/ml.

[1] The abbreviations used are: ELISA, enzyme-linked immunosorbent assay; PBS, phosphate-buffered saline; PCR, polymerase chain reaction.

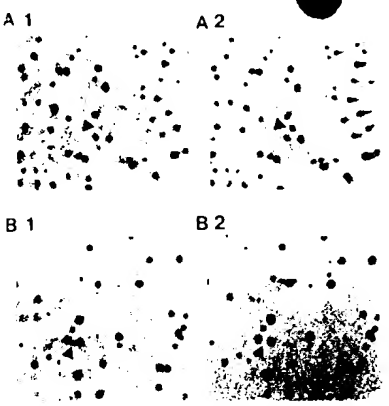


FIG. 4. **Testing of cross-reaction of proteins in plaques obtained from libraries of mutagenized DNA.** The results of two experiments are shown as examples (*A* and *B*). *A1* and *B1* show cross-reaction with ARM191, and *A2* and *B2* that with ARM193. The procedure was described in detail in the legend to Fig. 1. *Arrowheads* indicate mutants which showed altered cross-reaction; mutant D (the top ones in *A1* and *A2*) shows no cross-reaction with ARM191, and mutant 47 (*B1* and *B2*) none with ARM193.

TABLE I

Isolation of mutants by region-specified PCR mutagenesis

dITP	Total plaques expressing recA genes	Number of plaques picked up on the first selection	Total number of mutants identified	Number of species of mutants
μM				
0	1800	5	1	1
0.2	1100	21	4	2
20	2200	5	3	3
200	2000	21	11	10

Testing of the Competition between ARM191 and ARM193—The IgGs, ARM191 and ARM193, belong to subclasses 1 and 2b, respectively (Ikeda *et al.*, 1990), and thus each could be assayed by use of anti-mouse IgG_1 and anti-mouse IgG_{2b} antibodies, respectively. A solution of the purified recA protein (0.2 μg/ml) was put into the wells of a microtiter plate to coat the walls of the wells. Each solution of an indicated amount (no dilution: 1.5 μg/ml) of a tested anti-recA protein IgG (50 μl) contained 0 or 5 μg of the other anti-recA protein IgG/ml. The amounts of the tested anti-recA protein IgG that bound to the wells were measured by ELISA with anti-mouse IgG_1 or anti-mouse IgG_{2b} antibodies (Zymed Laboratories, Inc.).

Region-specified PCR Mutagenesis—An outline of the procedure is given in Fig. 1. pBEU14 DNA, which contains the *Escherichia coli recA* gene (Uhlin and Clark, 1981), was linearized with *Bam*HI and used as the template for PCR (polymerase chain reaction). A DNA region encoding the *recA* gene flanked by primer 1 (5′-ATGGCTATCGACGAAAACAA-3′) and primer 2 (5′-GAATTCTGTCATGGCATATCCTT-3′) was amplified by 25 cycles of PCR (see Fig. 1). Unless otherwise stated, the reaction mixture for PCR comprised 1 μM each of primers 1 and 2, about 3 pM the template DNA, 200 μM each of dATP, dTTP, dGTP, and dCTP, 200 μM deoxyinosine 5′-triphosphate (dITP), 0.025 units of *Taq* DNA polymerase/μl, 1.5 mM $MgCl_2$, 50 mM KCl, 0.001% gelatin, and 10 mM Tris-HCl buffer (pH 8.3). Each cycle of PCR consisted of: (i) incubation at 37 °C for 2 min, for loading of the primers onto the template DNA, (ii) incubation at 72 °C for 3 min, for polymerization, and (iii) incubation at 94 °C for 1 min, for denaturation. The amplified DNA was treated with *Eco*RI and the *Eco*RI fragment encoding the C-terminal 93-amino acid region was isolated by gel electrophoresis, followed by trapping on a DEAE membrane (NA45, Schleicher & Schuell). Then, DNA libraries containing the mutagenized *Eco*RI fragments were constructed on the λgt11 expression vector by ligating them at an *Eco*RI site of the vector. With appropriate orientation of a fragment relative to the vector, the C-terminal region of the *recA* gene was connected to the *lacZ* gene in-frame. Plaques of the phages in the libraries were obtained with the *E. coli* Y1090 strain as a host. Proteins in the plaques were transferred to a pair of membranes and then the cross-reaction with either ARM191 or ARM193 anti-recA protein IgG was examined by means of immunoblotting experiments, as described previously (Ikeda *et al.*, 1990; Morishima *et al.*, 1990). The plaques showing cross-reaction with only one of the IgGs (indicated by *arrows* in Fig. 1; examples are shown in Fig. 4) were picked up, and the phages were obtained after repeated single plaque isolation and testing by means of immunoblotting experiments.

DNA Sequence Analysis—The tested DNAs were cloned on pUC119 and then their sequences were analyzed by the dideoxyribonucleotide chain termination method (Sanger *et al.*, 1977); the labeling reaction was carried out by use of the double-stranded template according to a manual for Sequenase (United States Biochemical Co., Cleveland, OH), and the products were analyzed with the use of an automated DNA sequence analyzer (Du Pont GENESIS2000). We analyzed both strands in most of the cases.

Expression of Mutant recA Genes and Preparation of Cell-free Extracts—*Eco*RI fragments which carried mutation(s) in the C-terminal 93-amino acid region were ligated to the *Eco*RI site of the DNA region encoding the N-terminal 260 amino acids, which was under the control of the Tac promoter on a multicopy plasmid, pMI996 (Fig. 2), a derivative of pKK223-3 (Brosius and Holy, 1984). An *E. coli* strain, MV1184 (deletion of the *srl-recA* locus) (Vieira and Messing, 1987), was transformed with the DNA. The transformants were grown at 37 °C to the mid-logarithmic growth phase in 3-ml cultures and then expression of the mutated *recA* genes was induced by isopropyl-β-D-thiogalactoside treatment (0.2 mg/ml) for 2 h. After the treatment, the cells were collected by centrifugation and suspended in 200 μl of a lysis buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.1 M NaCl, and 5% Triton X-100), and then treated with lysozyme (0.6 mg/ml) at 0 °C for 15 min, followed by the addition of KCl (at 0.24 M). The lysate was centrifuged at about 15,000 × *g* for 15 min and the supernatant of the lysate was saved. The supernatant was diluted about 20-fold with PBS and then subjected to ELISA as described above. In the case of mutants 2 and 38, the precipitates of the lysates were resuspended in 5 M urea containing 100 mM Tris-HCl (pH 7.9), followed by centrifugation, and the mutant proteins were extracted from the precipitates with 100 mM Tris-HCl buffer (pH 7.9) containing 8 M urea and 0.1 M NaCl at 25 °C for 1 h.

RESULTS

Testing of the Competition between ARM191 and ARM193—The preliminary mapping of the epitopes of ARM191 and ARM193 indicated that both epitopes were located between Phe^{260} and Glu^{347} (Ikeda *et al.*, 1990). First, we examined whether or not these anti-recA protein-IgGs showed competition in the cross-reaction with the recA protein. Since ARM191 and ARM193 belong to different IgG subclasses, each can be discriminated through the use of subclass-specific antibodies on ELISA. Fig. 3 indicates that the presence of one of these IgGs did not affect the binding of the other IgG to the recA protein. We conclude that the epitope for ARM191 and ARM193 are different. Thus, we tried to locate the epitopes of these anti-recA protein IgGs more precisely.

Development of Region-specified PCR Mutagenesis and Isolation of Mutants—Since we were not able to locate the epitopes of ARM191 and ARM193 by examinations of the cross-reaction of subfragments of the recA polypeptide, we developed a novel technique for region-specified mutagenesis which could be applied for mapping of the epitopes. The whole process for the isolation of mutants causing altered cross-reaction consists of three stages (Fig. 1); the introduction of region-specified random base substitutions by means of PCR, construction of libraries of the mutagenized DNA with the use of the λgt11 expression vector, and *in situ* testing for cross-reaction of the mutated polypeptides expressed in the plaques obtained from the libraries. In PCR (Saiki *et al.*, 1985), *Taq* DNA polymerase causes the misincorporation of nucleotides (Eckert and Kunkel, 1990). We added deoxyinosine 5′-triphosphate (dITP), at 200 μM, to the reaction mixture for PCR to enhance the misincorporation (Martin and Castro, 1985), and specifically amplified a DNA region defined by a pair of oligonucleotides (primers 1 and 2). Primer 1 includes the initiation codon of the recA polypeptide and

TABLE II

Mutations affecting cross-reaction with anti-recA protein IgG ARM191 or ARM193

Mutation site on recA polypeptide	Name of mutant	Change in DNA sequence	Amino acid replaced	Cross reactivity on immunoblotting[a]		Half-maximum values on ELISA			
				ARM191	ARM193	ARM191	ARM193	ARM414	MAb156
						μg/ml			
Wild type				+	+				
Experiment 1						0.017	0.27	0.12	0.019
Experiment 2						0.085	0.21	0.19	0.017
Experiment 3						0.021	0.75	0.12	0.012
Average of "plus" recA polypeptide (S.D.)						0.023	0.28	0.14	0.015
						(0.024)	(0.18)	(0.11)	(0.004)
283	38	C[c]G → C[c]G	Leu → Pro	−	+	0.13	0.17	0.075	>6.0
291	D	T[a]C → T[c]C	Tyr → Cys	−	+	>30	0.27	0.24	>30
296	43	G[a]G → G[c]G	Glu → Gly	−	+	>30	0.53	0.19	>30
305	4	G[c]G → G[t]G	Ala → Glx	−	+	0.75	0.30	0.21	1.7
308	32	TG[c] → TG[t]	Trp → Cys	−	+	>30	0.13	0.21	>30
309	2, 5, 12	C[c]G → C[t]G	Leu → Pro	−	+	>7.5	0.15	0.053	>30
312	37, 39	AA[c] → AA[a]	Asn → Lys	−	+	>9.5	0.19	0.11	>9.5
315	30.1[d]	A[c]C → A[t]C	Thr → Ile	+	−	0.013	>10	0.052	0.015
332	31, 33.1[c,d]	[a]AC → [g]AC	Asn → Asp	+	−	0.012	>10	0.06	0.013
333	23, 34	[t]CA → [c]CA	Ser → Pro	+	−	0.015	>10	0.04	0.019
337	47	[t]TC → [c]TC	Phe → Ser	+	−	NT[e]	NT	NT	NT
338	40	T[c]T → T[t]T	Ser → Phe	+	−	0.0095	>30	0.075	0.015
294	36.1	A[a]A → A[g]A	Lys → Arg	+	−	0.013	>10	0.075	0.0075
330	36.2	[a]AC → [g]AC	Asn → Asp						
319	41.1[d]	A[c]C → A[t]C	Ile → Thr	−	+	>30	0.085	0.48	>30
323	41.2	G[t]A → G[c]A	Val → Ala						
267	30.2, 33.2, 36.3	GG[t] → GG[c]	No replacement						
321	30.3	AA[c] → AA[a]	No replacement						

[a] Proteins showing the plus phenotype with respect to cross-reaction with the relevant IgG.

[b] Shadow letters indicate bases substituted.

[c] All mutants having the same substitution were isolated from the same PCR preparation, except mutants 23 and 34.

[d] Mutants 30, 33, 36, and 41 have more than two base substitutions, each of which is indicated, *e.g.* 30.1 and 30.2.

[e] NT, not tested.

primer 2 includes a complementary sequence about 80 bases downstream of its stop codon (Fig. 1). Primer 2 was designed so as to have an *Eco*RI cutting site. At more than 200 μM, dITP severely inhibited the amplification of DNA by PCR (data not shown). After 25 cycles of PCR, the amplified DNA was treated with *Eco*RI and the *Eco*RI fragment encoding the region of the recA polypeptide extending from Phe^{200} to the C-terminal Phe was isolated. The fragment was inserted into the *Eco*RI site of the λgt11 expression vector (Young and Davis, 1983). The proteins in the plaques were examined as to their *in situ* cross-reaction with anti-recA protein IgGs ARM191 and ARM193. As shown in Fig. 4, most of the plaques expressing the fused recA polypeptide showed cross-reaction to both IgGs. We could pick up about 1% of these plaques as candidates of mutants which showed cross-reaction to only one of the IgGs (indicated by *arrowheads* in Fig. 4; Table I).

From the selected plaques, phages were recovered and the inserted *Eco*RI fragments were reisolated. The DNA fragments were recloned on the pUC119 vector (Vieira and Messing, 1987) and their DNA sequences were analyzed by the dideoxyribonucleotide chain termination method (Sanger *et al.*, 1977). We analyzed both strands in most of the cases by use of double-stranded form and an appropriate primer. The triplets which were changed by mutations are shown in Table II. During the development of the new technique for mutagenesis, we obtained 19 mutants exhibiting altered cross-reaction to either ARM191 or ARM193 from among the 52 plaques tested (Table I). So far as defined, all mutations were the substitution of amino acid(s) caused by base substitution(s) (Table II). Some of the mutants have a second (third) base substitution which does not affect the amino acid sequence. Of the base substitutions defined, 83% are transitions and 17% transversions.

On the first screening, the addition of dITP appeared to have only a little effect on the PCR mutagenesis (Table I). However, when we characterized the mutants isolated, we found that the addition of dITP is essential for this mutagenesis (Table I). Without dITP or with 0.2 μM dITP, we picked up 5 and 21 plaques, respectively, on the first screening, but obtained only 3 kinds of mutants. When dITP was added at 200 μM, we picked up 21 plaques from among about 2000 plaques expressing the fused recA polypeptide, and from these groups we finally isolated 11 mutants which were categorized into 10 kinds. Three of them (mutants 30, 33, and 36) had base substitutions at two or three sites (see Table II). These findings indicate the significant enhancement of the yield and variation on mutagenesis with dITP, and the very high yield of base substitutions under an appropriate set of conditions.

Some mutants have base substitutions at the same sites. Except mutants 23 and 34 (substitution of Ser^{333}), such overlapped mutants were isolated from the same PCR preparations. Thus, these overlapped mutations seem to be created at an early cycle in PCR and amplified during the procedure

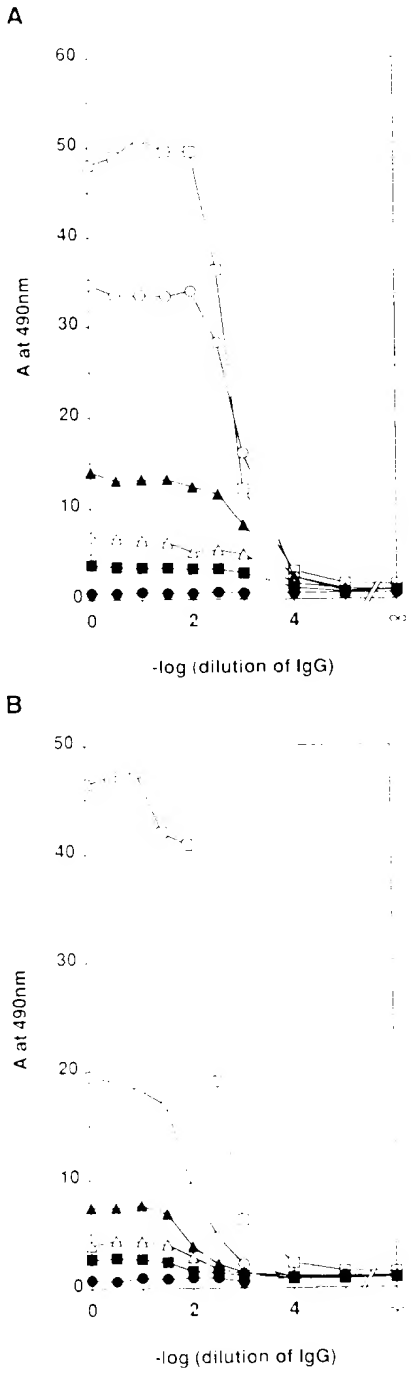


FIG. 5. **Variation of the antigen concentration on ELISA does not change the amount of IgG giving a half-maximum signal.** The wells of a microtiter plate were coated with the purified recA protein at the indicated concentrations. On the other hand, the concentration of the anti-recA protein IgG (ARM191 in *A*; ARM193 in *B*) was adjusted to 30 μg/ml ("nondilution"), and then a series of dilutions of the IgG was put into individual wells. The bound IgG was measured by ELISA. The concentrations of the recA protein were as follows: □, 2 μg/ml; ○, 0.2 μg/ml; ▲, 0.1 μg/ml; △, 0.05 μg/ml; ■, 0.02 μg/ml; ●, without recA protein.

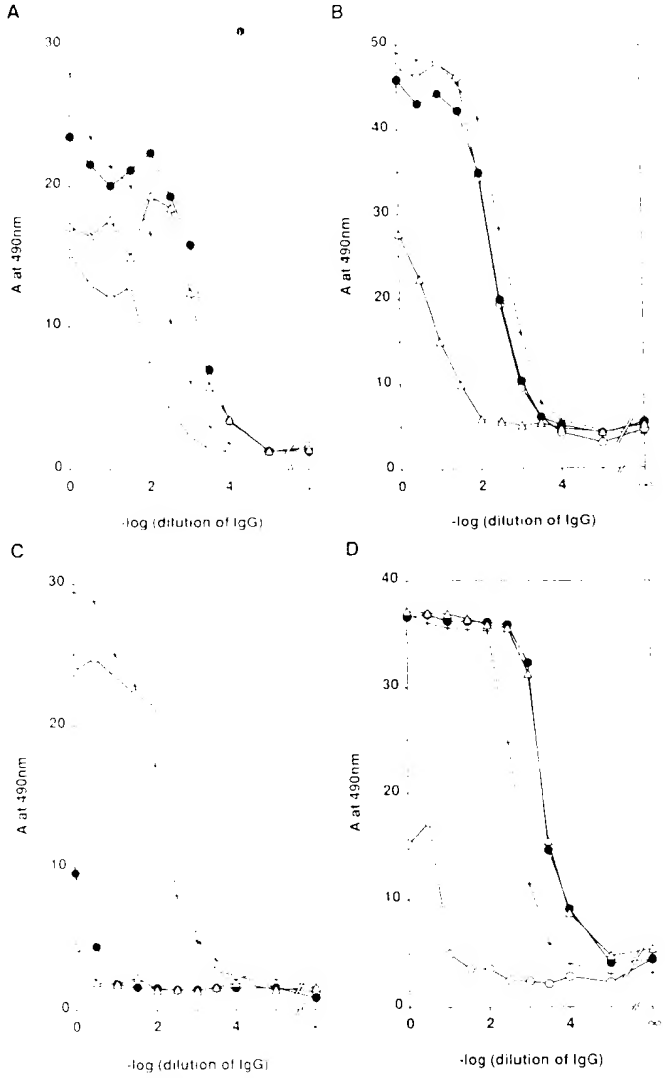


FIG. 6. **Examples of ELISA for testing the cross-reaction of mutant recA proteins to anti-recA protein IgGs, ARM191, ARM193, and MAb156.** Cell-free extracts were prepared from cells in which the mutant *recA* genes were expressed; *A*, wild type; *B*, mutant 38; *C*, mutant 32; *D*, mutant 31. The cell-free extracts were put into individual wells of a microtiter plate to allow the adsorption of proteins. Then, the cross-reaction to each of the IgGs was examined by ELISA, as described in the legend to Fig. 5. The IgGs used were: ●, ARM191; ○, ARM193; △, MAb156. Anti-recA protein IgG, ARM414, of which epitopes is located between Glu^{233} and Lys^{256} (Ikeda *et al.*, 1990) was used as a positive control (+).

Except for in these cases, base substitutions appeared to be introduced at random in the amplified DNA.

Cross-reaction of Mutant recA Proteins—We constructed a plasmid (pMI996; Fig. 2) in which the wild type *recA* gene was under the control of the *Tac* promoter on a multicopy plasmid (a derivative of pKK223-3 (Brosius and Holy, 1984). We replaced the *Eco*RI-*Eco*RI region encoding the C-terminal 93 amino acids of the wild type recA polypeptide with the *Eco*RI fragments on which we found mutation(s) (Fig. 2). Then, a mutant of *E. coli* in which the whole *recA* gene was deleted was transformed with these plasmids. The expression of the mutant *recA* genes was induced and cell-free extracts were prepared, followed by quantitative assaying (ELISA) for cross-reaction with ARM191 and ARM193. For comparison of the extent of cross-reaction of a tested IgG with the recA protein, we determined the amount of the IgG giving a half-maximum signal on ELISA. As shown in Fig. 5, variations in the amount of recA protein did not significantly change the amount of the IgG which gave the half-maximum signal on ELISA. We calculated the amounts of ARM191 and ARM193 giving a half-maximum signal to be 0.0277 (σ = 0.017) and 0.29 (σ = 0.10) μg/ml, respectively, from the data in this figure.

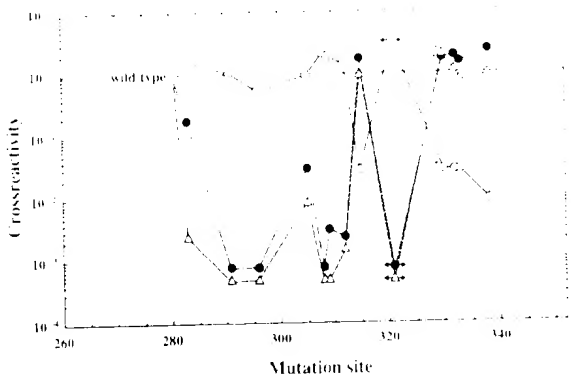


FIG. 7. **Mutation sites and alterations in cross-reactivity.** The *numbers* under the *abscissa* are the amino acid positions from the N terminus of the recA polypeptide. Cross-reactivity is defined as follows: (the amount of an IgG giving the half-maximum value on ELISA with wild-type recA protein)/(the amount of an IgG giving the half-maximum value on ELISA with mutant recA protein). Thus, it shows the deficiency in the cross-reaction. ●, cross reaction with ARM191; ○, that with ARM193; △, that with MAb156. *Symbols* within the *shadowed area* indicate that no cross reaction was detected (within the limit of the measurement, which is indicated by each *symbol*). The *symbols* with a *horizontal arrow* around position 320 are those for a mutant having two amino acid substitutions at positions 319 and 323, and those with * at position 330 are those for a mutant having two amino acid substitutions at positions 294 and 330.

Examples of ELISA with mutant recA proteins as well as the wild type protein are shown in Fig. 6. The values calculated on ELISA are listed in Table II. In order to obtain the half-maximum signals on ELISA, all proteins (except mutant 38) which had been determined to be mutants by immunoblotting experiments were shown to require at least 10-fold more IgG ARM191 or ARM193 than the average for plus proteins (Fig. 7). The results shown in Fig. 7 clearly indicate that the mutants affecting the cross-reaction with ARM191 and those affecting that with ARM193 are separately clustered, and only slightly overlap each other (between positions 315 and about 320). This conclusion is consistent with the absence of competition in the binding of these IgGs to the recA protein (Fig. 3).

Comparison of MAb156 with ARM191 and ARM193—MAb156 was isolated by Karu and Allen (Karu and Allen, 1982). Since it was assumed that the epitope of this antibody is located near the C terminus of the recA polypeptide,* we examined the cross-reaction of MAb156 with the mutant recA proteins isolated in the present study. All of the mutations affecting the cross-reaction with ARM191 also affected the cross-reaction with MAb156, but none of the mutations affecting the cross-reaction with ARM193 did. This suggests that the epitope for MAb156 and that of ARM191 are similar, but the following results show that they are not identical. Most of the mutations abolishing the cross-reaction with ARM191 also abolished that with MAb156 (Fig. 7), but substitution of Leu^{283} by Pro in mutant 38 strongly interfered with the cross-reaction with MAb156, but not so much with that with ARM191 (Figs. 6*B* and 7). The difference in the mode of cross-reaction between ARM191 and MAb156 was also observed with a mutant recA protein having a substitution of Ala^{305} by Gly (Fig. 7).

DISCUSSION

Region-specific PCR mutagenesis is an efficient tool for introducing random base substitution mutations specifically in a region defined by a pair of primers. From among the plaques expressing the fused *recA* gene, we obtained mutants of the *recA* gene at the frequency of 0.5% under the optimized conditions. Using this mutagenesis, we identified regions of the recA polypeptide in which amino acid substitutions prevent the cross-reaction with ARM191, ARM193, and/or MAb156. The region for ARM191 and that for ARM193 are different, but slightly overlap each other; *i.e.* that for the former IgG comprises positions 283 through about 320, and that for the latter positions through 315–338. Since ARM191 and ARM193 showed no competition, Thr^{315}, and Ile^{319} and Ile^{319} and/or Val^{323} would be located on different sides of a local structure or the whole molecule of the recA polypeptide.

The substitution of an amino acid might interfere with the cross-reaction with an IgG directly or through a very local change in the tertiary structure, or indirectly through extensive alteration of the tertiary structure of the recA protein. The Leu^{283} to Pro substitution in mutant 38 could cause gross alteration of the tertiary structure of the recA protein. The clustering of other amino acid substitutions in a particular region is favorable for the first two possibilities rather than the last one. ARM193 was shown to prevent the self-assembly of the recA protein and ARM191 to inhibit the binding of double-stranded DNA. Thus, the active sites involved in these functions would be located in the space at or around the relevant epitope. We are testing these possibilities by examining the effects of the mutant recA proteins isolated in this study and those constructed by another round of region-specified PCR mutagenesis. Determination of the epitope loci and the inhibitory effects of these IgGs, together with the tertiary structure of the recA protein, will facilitate the understanding of the function of the protein in relation to its structure.

The technique of region-specified random base substitutions involving the use of PCR employed in this study is very useful not only for epitope mapping, as described in this paper, but is also widely useful for studies on the function of a gene and an enzyme or protein, because of the flexibility as to specifying a target region, and the high yield of the random base substitutions.

Acknowledgments—We thank Dr. Alexander Karu (University of California, Berkeley) and Dr. A. John Clark (University of California, Berkeley) for sending us the purified anti-recA protein monoclonal IgG, MAb156, and the related information before publication.

REFERENCES

Berger, S. L., and Kimmel, A. R. (1987) *Guide to Molecular Cloning Techniques Methods in Enzymology*, Vol. 152, Academic, Orlando
Brenner, J., and Holy, A. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6929–6933
Eckert, K., and Kunkel, T. A. (1990) *Nucleic Acid Res.* **18**, 3739–3744
Ikawa, S., Kamiya, N., and Shibata, T. (1989) *J. Biol. Chem.* **264**, 21167–21176
Ikeda, M., Makino, O., and Shibata, T. (1990) *J. Biol. Chem.* **265**, 8948–8956
Karu, A. E., and Allen, L. (1982) *Biochimie (Paris)* **64**, 723–729
Makino, O., Shibata, Y., Maeda, H., Shibata, T., and Ando, T. (1985) *J. Biol. Chem.* **260**, 15402–15405
Makino, O., Ikawa, S., Shibata, Y., Maeda, H., Ando, T., and Shibata, T. (1987) *J. Biol. Chem.* **262**, 12237–12246
Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Martin, F. H., and Castro, M. M. (1985) *Nucleic Acid Res.* **13**, 8927–8938
Morishima, N., Nakagawa, K., Yamamoto, E., and Shibata, T. (1990) *J. Biol. Chem.* **265**, 15189–15197
Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) *Science* **230**, 1350–1354
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
Shibata, T., Cunningham, R. P., and Radding, C. M. (1981) *J. Biol. Chem.* **256**, 7557–7564
Shibata, T., Ikeda, M., Makino, O., and Ikawa, S. (1991) *Biochimie (Paris)* **73**, 209–218
Uhlin, B. E., and Clark, A. J. (1981) *J. Bacteriol.* **148**, 386–390
Vieira, J., and Messing, J. (1987) *Methods Enzymol.* **153**, 3–11
Young, R. A., and Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1194–1198

* A. J. Clark and A. Karu, personal communication.

March 25, 1992 VOLUME 267 NUMBER 9

ISSN 0021-9258
JBCHA3 267(9) 5723-6450 (1992)

THE Journal of Biological Chemistry

Published by the American Society for Biochemistry and Molecular Biology

FOUNDED BY CHRISTIAN A. HERTER

AND SUSTAINED IN PART BY THE CHRISTIAN A. HERTER MEMORIAL FUND